

# Proof of the Ergodic Hypothesis for Typical Hard Ball Systems

Nándor Simányi\*

**Abstract.** We consider the system of  $N (\geq 2)$  hard balls with masses  $m_1, \dots, m_N$  and radius  $r$  in the flat torus  $\mathbb{T}_L^\nu = \mathbb{R}^\nu / L \cdot \mathbb{Z}^\nu$  of size  $L$ ,  $\nu \geq 3$ . We prove the ergodicity (actually, the Bernoulli mixing property) of such systems for almost every selection  $(m_1, \dots, m_N; L)$  of the outer geometric parameters. This theorem complements my earlier result that proved the same, almost sure ergodicity for the case  $\nu = 2$ . The method of that proof was primarily dynamical-geometric, whereas the present approach is inherently algebraic.

## 1 Introduction

Hard ball systems or, a bit more generally, mathematical billiards constitute an important and quite interesting family of dynamical systems being intensively studied by dynamicists and researchers of mathematical physics, as well. These dynamical systems pose many challenging mathematical questions, most of them concerning the ergodic (mixing) properties of such systems. The introduction of hard ball systems and the first major steps in their investigations date back to the 40's and 60's, see Krylov's paper [K(1942)] and Sinai's ground-breaking works [Sin(1963)] and [Sin(1970)], in which the author – among other things – formulated the modern version of Boltzmann's ergodic hypothesis (what we call today the Boltzmann-Sinai ergodic hypothesis) by claiming that every hard ball system in a flat torus is ergodic, of course after fixing the values of the trivial flow-invariant quantities. In the articles [Sin(1970)] and [B-S(1973)] Bunimovich and Sinai proved this hypothesis for two hard disks on the two-dimensional unit torus  $\mathbb{T}^2$ . The generalization of this result to higher dimensions  $\nu > 2$  took fourteen years, and was done by Chernov and Sinai in [S-Ch(1987)]. Although the model of two hard balls in  $\mathbb{T}^\nu$  is already rather involved technically, it is still a so-called strictly dispersive billiard system, i.e., such that the smooth components of the boundary  $\partial \mathbf{Q}$  of the configuration space are strictly concave from outside  $\mathbf{Q}$ . (They are bending away from  $\mathbf{Q}$ .) The billiard systems of more than two hard balls in  $\mathbb{T}^\nu$  are no longer strictly dispersive, but just semi-dispersive (strict concavity of the smooth components of  $\partial \mathbf{Q}$  is lost, merely concavity persists), and this circumstance causes a lot of additional technical troubles in their study. In the series of my

---

\*Research supported by the National Science Foundation, grant DMS-0098773.

joint papers with A. Krámli and D. Szász [K-S-Sz(1989)], [K-S-Sz(1990)], [K-S-Sz(1991)], and [K-S-Sz(1992)] we developed several new methods, and proved the ergodicity of more and more complicated semi-dispersive billiards culminating in the proof of ergodicity of four billiard balls in the torus  $\mathbb{T}^\nu$  ( $\nu \geq 3$ ), [K-S-Sz(1992)]. Then, in 1992, Bunimovich, Liverani, Pellegrinotti and Sukhov [B-L-P-S(1992)] were able to prove the ergodicity for some systems with an arbitrarily large number of hard balls. The shortcoming of their model, however, is that, on one hand, they restrict the types of all feasible ball-to-ball collisions, on the other hand they introduce some additional scattering effect with the collisions at the strictly concave wall of the container. The only result with an arbitrarily large number of balls in a flat unit torus  $\mathbb{T}^\nu$  was achieved in the twin papers of mine [Sim(1992-I-II)], where I managed to prove the ergodicity (actually, the K-mixing property) of  $N$  hard balls in  $\mathbb{T}^\nu$ , provided that  $N \leq \nu$ . The annoying shortcoming of that result is that the larger the number of balls  $N$  is, larger and larger dimension  $\nu$  of the ambient container is required by the method of the proof.

On the other hand, if someone considers a hard ball system in an elongated torus which is long in one direction but narrow in the others, so that the balls must keep their cyclic order in the “long direction” (Sinai’s “pen-case” model), then the technical difficulties can be handled, thanks to the fact that the collisions of balls are now restricted to neighboring pairs. The hyperbolicity of such models in three dimensions and the ergodicity in dimension four have been proved in [S-Sz(1995)].

The positivity of the metric entropy for several systems of hard balls can be proven relatively easily, as was shown in the paper [W(1988)]. The articles [L-W(1995)] and [W(1990)] are nice surveys describing a general setup leading to the technical problems treated in a series of research papers. For a comprehensive survey of the results and open problems in this field, see [Sz(1996)].

Pesin’s theory [P(1977)] on the ergodic properties of non-uniformly hyperbolic, smooth dynamical systems has been generalized substantially to dynamical systems with singularities (and with a relatively mild behavior near the singularities) by A. Katok and J-M. Strelcyn [K-S(1986)]. Since then, the so-called Pesin’s and Katok-Strelcyn’s theories have become part of the folklore in the theory of dynamical systems. They claim that – under some mild regularity conditions, particularly near the singularities – every non-uniformly hyperbolic and ergodic flow enjoys the Kolmogorov-mixing property, shortly the K-mixing property.

Later on it was discovered and proven in [C-H(1996)] and [O-W(1998)] that the above-mentioned fully hyperbolic and ergodic flows with singularities turn out to be automatically having the Bernoulli mixing (B-mixing) property. It is worth noting here that almost every semi-dispersive billiard system, especially every hard ball system, enjoys those mild regularity conditions imposed on the systems (as axioms) by [K-S(1986)], [C-H(1996)], and [O-W(1998)]. In other words, for a hard ball flow  $(\mathbf{M}, \{S^t\}, \mu)$  the (global) ergodicity of the system actually implies its full hyperbolicity and the B-mixing property, as well.

Finally, in our joint venture with D. Szász [S-Sz(1999)], we prevailed over the difficulty caused by the low value of the dimension  $\nu$  by developing a brand

new algebraic approach for the study of hard ball systems. That result, however, only establishes complete hyperbolicity (nonzero Lyapunov exponents almost everywhere) for  $N$  balls in  $\mathbb{T}^\nu$ . The ergodicity appeared to be a harder task.

We note, however, that the algebraic method developed in [S-Sz(1999)] is being further developed in this paper in order to obtain ergodicity, not only full hyperbolicity.

Consider the  $\nu$ -dimensional ( $\nu \geq 2$ ), standard, flat torus  $\mathbb{T}_L^\nu = \mathbb{R}^\nu / L \cdot \mathbb{Z}^\nu$  as the vessel containing  $N$  ( $\geq 2$ ) hard balls (spheres)  $B_1, \dots, B_N$  with positive masses  $m_1, \dots, m_N$  and (just for simplicity) common radius  $r > 0$ . We always assume that the radius  $r > 0$  is not too big, so that even the interior of the arising configuration space  $\mathbf{Q}$  is connected. Denote the center of the ball  $B_i$  by  $q_i \in \mathbb{T}^\nu$ , and let  $v_i = \dot{q}_i$  be the velocity of the  $i$ -th particle. We investigate the uniform motion of the balls  $B_1, \dots, B_N$  inside the container  $\mathbb{T}^\nu$  with half a unit of total kinetic energy:  $E = \frac{1}{2} \sum_{i=1}^N m_i \|v_i\|^2 = \frac{1}{2}$ . We assume that the collisions between balls are perfectly elastic. Since – beside the kinetic energy  $E$  – the total momentum  $I = \sum_{i=1}^N m_i v_i \in \mathbb{R}^\nu$  is also a trivial first integral of the motion, we make the standard reduction  $I = 0$ . Due to the apparent translation invariance of the arising dynamical system, we factorize the configuration space with respect to uniform spatial translations as follows:  $(q_1, \dots, q_N) \sim (q_1 + a, \dots, q_N + a)$  for all translation vectors  $a \in \mathbb{T}^\nu$ . The configuration space  $\mathbf{Q}$  of the arising flow is then the factor torus  $\left( (\mathbb{T}^\nu)^N / \sim \right) \cong \mathbb{T}^{\nu(N-1)}$  minus the cylinders

$$C_{i,j} = \left\{ (q_1, \dots, q_N) \in \mathbb{T}^{\nu(N-1)} : \text{dist}(q_i, q_j) < 2r \right\}$$

( $1 \leq i < j \leq N$ ) corresponding to the forbidden overlap between the  $i$ -th and  $j$ -th spheres. Then it is easy to see that the compound configuration point

$$q = (q_1, \dots, q_N) \in \mathbf{Q} = \mathbb{T}^{\nu(N-1)} \setminus \bigcup_{1 \leq i < j \leq N} C_{i,j}$$

moves in  $\mathbf{Q}$  uniformly with unit speed and bounces back from the boundaries  $\partial C_{i,j}$  of the cylinders  $C_{i,j}$  according to the classical law of geometric optics: the angle of reflection equals the angle of incidence. More precisely: the post-collision velocity  $v^+$  can be obtained from the pre-collision velocity  $v^-$  by the orthogonal reflection across the tangent hyperplane of the boundary  $\partial \mathbf{Q}$  at the point of collision. Here we must emphasize that the phrase “orthogonal” should be understood with respect to the natural Riemannian metric (the so-called mass metric)  $\|dq\|^2 = \sum_{i=1}^N m_i \|dq_i\|^2$  in the configuration space  $\mathbf{Q}$ . For the normalized Liouville measure  $\mu$  of the arising flow  $\{S^t\}$  we obviously have  $d\mu = \text{const} \cdot dq \cdot dv$ , where  $dq$  is the Riemannian volume in  $\mathbf{Q}$  induced by the above metric and  $dv$  is the surface measure (determined by the restriction of the Riemannian metric above) on the

sphere of compound velocities

$$\mathbb{S}^{\nu(N-1)-1} = \left\{ (v_1, \dots, v_N) \in (\mathbb{R}^\nu)^N : \sum_{i=1}^N m_i v_i = 0 \text{ and } \sum_{i=1}^N m_i \|v_i\|^2 = 1 \right\}.$$

The phase space  $\mathbf{M}$  of the flow  $\{S^t\}$  is the unit tangent bundle  $\mathbf{Q} \times \mathbb{S}^{d-1}$  of the configuration space  $\mathbf{Q}$ . (We will always use the shorthand notation  $d = \nu(N-1)$  for the dimension of the billiard table  $\mathbf{Q}$ .) We must, however, note here that at the boundary  $\partial\mathbf{Q}$  of  $\mathbf{Q}$  one has to glue together the pre-collision and post-collision velocities in order to form the phase space  $\mathbf{M}$ , so  $\mathbf{M}$  is equal to the unit tangent bundle  $\mathbf{Q} \times \mathbb{S}^{d-1}$  modulo this identification.

A bit more detailed definition of hard ball systems with arbitrary masses, as well as their role in the family of cylindric billiards, can be found in §4 of [S-Sz(2000)] and in §1 of [S-Sz(1999)]. We denote the arising flow by  $(\mathbf{M}, \{S^t\}_{t \in \mathbb{R}}, \mu)$ .

In the series of articles [K-S-Sz(1989)], [K-S-Sz(1991)], [K-S-Sz(1992)], [Sim(1992-I)], and [Sim(1992-II)] the authors developed a powerful, three-step strategy for proving the (hyperbolic) ergodicity of hard ball systems. First of all, all these proofs are inductions on the number  $N$  of balls involved in the problem. Secondly, the induction step itself consists of the following three major steps:

**Step I.** To prove that every non-singular (i.e., smooth) trajectory segment  $S^{[a,b]}x_0$  with a “combinatorially rich” (in a well-defined sense) symbolic collision sequence is automatically sufficient (or, in other words, “geometrically hyperbolic”, see below in §2), provided that the phase point  $x_0$  does not belong to a countable union  $J$  of smooth sub-manifolds with codimension at least two. (Containing the exceptional phase points.)

The exceptional set  $J$  featuring this result is negligible in our dynamical considerations – it is a so-called slim set. For the basic properties of slim sets, see §2 below.

**Step II.** Assume the induction hypothesis, i.e., that all hard ball systems with  $N'$  balls ( $2 \leq N' < N$ ) are (hyperbolic and) ergodic. Prove that then there exists a slim set  $S \subset \mathbf{M}$  (see §2) with the following property: For every phase point  $x_0 \in \mathbf{M} \setminus S$  the entire trajectory  $S^{\mathbb{R}}x_0$  contains at most one singularity and its symbolic collision sequence is combinatorially rich, just as required by the result of Step I.

**Step III.** By using again the induction hypothesis, prove that almost every singular trajectory is sufficient in the time interval  $(t_0, +\infty)$ , where  $t_0$  is the time moment of the singular reflection. (Here the phrase “almost every” refers to the volume defined by the induced Riemannian metric on the singularity manifolds.)

We note here that the almost sure sufficiency of the singular trajectories (featuring Step III) is an essential condition for the proof of the celebrated theorem on local ergodicity for algebraic semi-dispersive billiards proved by Bálint-Chernov-Szász-Tóth in [B-Ch-Sz-T (2002)]. Under this assumption the theorem of [B-Ch-Sz-T (2002)] states that in any algebraic semi-dispersive billiard system (i.e., in

a system such that the smooth components of the boundary  $\partial\mathbf{Q}$  are algebraic hypersurfaces) a suitable, open neighborhood  $U_0$  of any sufficient phase point  $x_0 \in \mathbf{M}$  (with at most one singularity on its trajectory) belongs to a single ergodic component of the billiard flow  $(\mathbf{M}, \{S^t\}_{t \in \mathbb{R}}, \mu)$ .

In an inductive proof of ergodicity, steps I and II together ensure that there exists an arc-wise connected set  $C \subset \mathbf{M}$  with full measure, such that every phase point  $x_0 \in C$  is sufficient with at most one singularity on its trajectory. Then the cited theorem on local ergodicity (now taking advantage of the result of Step III) states that for every phase point  $x_0 \in C$  an open neighborhood  $U_0$  of  $x_0$  belongs to one ergodic component of the flow. Finally, the connectedness of the set  $C$  and  $\mu(\mathbf{M} \setminus C) = 0$  easily imply that the flow  $(\mathbf{M}, \{S^t\}_{t \in \mathbb{R}}, \mu)$  (now with  $N$  balls) is indeed ergodic, and actually fully hyperbolic, as well.

The main result of this paper is the

**Theorem.** In the case  $\nu \geq 3$  for almost every selection  $(m_1, \dots, m_N; L)$  of the outer geometric parameters from the region  $m_i > 0$ ,  $L > L_0(r, \nu)$ , where the interior of the phase space is connected, it is true that the billiard flow  $(\mathbf{M}_{\vec{m}, L}, \{S^t\}, \mu_{\vec{m}, L})$  of the  $N$ -ball system is ergodic and completely hyperbolic. Then, following from the results of Chernov-Haskell [C-H(1996)] and Ornstein-Weiss [O-W(1998)], such a semi-dispersive billiard system actually enjoys the B-mixing property, as well.

**Remark 1.** We note that the main result of this paper and that of [Sim(2003)] nicely complement each other. They precisely assert the same, almost sure ergodicity of hard ball systems in the cases  $\nu \geq 3$  and  $\nu = 2$ , respectively. It should be noted, however, that the proof of [Sim(2003)] is primarily dynamical-geometric (except the verification of the Chernov-Sinai Ansatz), whereas the novel parts of the present proof are fundamentally algebraic.

**Remark 2.** The above inequality  $L > L_0(r, \nu)$  corresponds to physically relevant situations. Indeed, in the case  $L < L_0(r, \nu)$  the particles would not have enough room even to freely exchange positions.

The paper is organized as follows: §2 provides all necessary prerequisites and technical tools that will be required by the proof of the theorem. Based on the results obtained in [S-Sz(1999)], the subsequent Section §3 carries out Step I of the inductive strategy outlined above, but for the case when the outer geometric parameters  $(m_1, \dots, m_N; L)$  are incorporated in the algebraic process as variables. (Just as the positions and velocities of the particles!) Finally, the closing Section §4 utilizes a “Fubini type argument” by proving Step I for almost every (with respect to the Lebesgue measure of the  $(m_1, \dots, m_N; L)$ -space) hard ball system ( $N \geq 2$ ,  $\nu \geq 3$ ). This will finish the inductive proof of the theorem, for Steps II and III of the induction strategy are easy consequences of some earlier results.

## 2 Prerequisites

### 2.1 Cylindric billiards

Consider the  $d$ -dimensional ( $d \geq 2$ ) flat torus  $\mathbb{T}^d = \mathbb{R}^d / \mathcal{L}$  supplied with the usual Riemannian inner product  $\langle \cdot, \cdot \rangle$  inherited from the standard inner product of the universal covering space  $\mathbb{R}^d$ . Here  $\mathcal{L} \subset \mathbb{R}^d$  is assumed to be a lattice, i.e., a discrete subgroup of the additive group  $\mathbb{R}^d$  with  $\text{rank}(\mathcal{L}) = d$ . The reason why we want to allow general lattices, other than just the integer lattice  $\mathbb{Z}^d$ , is that otherwise the hard ball systems would not be covered. The geometry of the structure lattice  $\mathcal{L}$  in the case of a hard ball system is significantly different from the geometry of the standard lattice  $\mathbb{Z}^d$  in the standard Euklidian space  $\mathbb{R}^d$ , see later in this section.

The configuration space of a cylindric billiard is  $\mathbf{Q} = \mathbb{T}^d \setminus (C_1 \cup \dots \cup C_k)$ , where the cylindric scatterers  $C_i$  ( $i = 1, \dots, k$ ) are defined as follows.

Let  $A_i \subset \mathbb{R}^d$  be a so-called lattice subspace of  $\mathbb{R}^d$ , which means that  $\text{rank}(A_i \cap \mathcal{L}) = \dim A_i$ . In this case the factor  $A_i / (A_i \cap \mathcal{L})$  is a sub-torus in  $\mathbb{T}^d = \mathbb{R}^d / \mathcal{L}$  which will be taken as the generator of the cylinder  $C_i \subset \mathbb{T}^d$ ,  $i = 1, \dots, k$ . Denote by  $L_i = A_i^\perp$  the ortho-complement of  $A_i$  in  $\mathbb{R}^d$ . Throughout this paper we will always assume that  $\dim L_i \geq 2$ . Let, furthermore, the numbers  $r_i > 0$  (the radii of the spherical cylinders  $C_i$ ) and some translation vectors  $t_i \in \mathbb{T}^d = \mathbb{R}^d / \mathcal{L}$  be given. The translation vectors  $t_i$  play a role in positioning the cylinders  $C_i$  in the ambient torus  $\mathbb{T}^d$ . Set

$$C_i = \{x \in \mathbb{T}^d : \text{dist}(x - t_i, A_i / (A_i \cap \mathcal{L})) < r_i\}.$$

In order to avoid further unnecessary complications, we always assume that the interior of the configuration space  $\mathbf{Q} = \mathbb{T}^d \setminus (C_1 \cup \dots \cup C_k)$  is connected. The phase space  $\mathbf{M}$  of our cylindric billiard flow will be the unit tangent bundle of  $\mathbf{Q}$  (modulo the natural gluing at its boundary), i.e.,  $\mathbf{M} = \mathbf{Q} \times \mathbb{S}^{d-1}$ . (Here  $\mathbb{S}^{d-1}$  denotes the unit sphere of  $\mathbb{R}^d$ .)

The dynamical system  $(\mathbf{M}, \{S^t\}_{t \in \mathbb{R}}, \mu)$ , where  $S^t$  ( $t \in \mathbb{R}$ ) is the dynamics defined by the uniform motion inside the domain  $\mathbf{Q}$  and specular reflections at its boundary (at the scatterers), and  $\mu$  is the Liouville measure, is called a cylindric billiard flow.

We note that the cylindric billiards – defined above – belong to the wider class of so-called semi-dispersive billiards, which means that the smooth components  $\partial \mathbf{Q}_i$  of the boundary  $\partial \mathbf{Q}$  of the configuration space  $\mathbf{Q}$  are (not necessarily strictly) concave from outside of  $\mathbf{Q}$ , i.e., they are bending away from the interior of  $\mathbf{Q}$ . As to the notions and notations in connection with semi-dispersive billiards, the reader is kindly referred to the article [K-S-Sz(1990)].

Throughout this paper we will always assume – without explicitly stating – that the considered semi-dispersive billiard system fulfills the following conditions:

$$\text{int} \mathbf{Q} \text{ is connected, and} \tag{2.1.1}$$

$$\begin{aligned} &\text{the } d\text{-dim spatial angle } \alpha(q) \text{ subtended by } \mathbf{Q} \\ &\text{at any of its boundary points } q \in \partial \mathbf{Q} \text{ is uniformly positive.} \end{aligned} \tag{2.1.2}$$

We note, however, that in the case of hard ball systems with a fixed radius  $r$  of the balls (see below) the non-degeneracy condition (2.1.2) only excludes countably many values of the size  $L$  of the container torus  $\mathbb{T}_L^\nu = \mathbb{R}^\nu / L \cdot \mathbb{Z}^\nu$  from the region  $L > L_0(r, \nu)$  where (2.1.1) is true. Therefore, in the sense of our theorem of “almost sure ergodicity”, the non-degeneracy condition (2.1.2) does not mean a restriction of generality.

## 2.2 Hard ball systems

Hard ball systems in the flat torus  $\mathbb{T}_L^\nu = \mathbb{R}^\nu / L \cdot \mathbb{Z}^\nu$  ( $\nu \geq 2$ ) with positive masses  $m_1, \dots, m_N$  are described (for example) in §1 of [S-Sz(1999)]. These are the dynamical systems describing the motion of  $N$  ( $\geq 2$ ) hard balls with a common radius  $r > 0$  and positive masses  $m_1, \dots, m_N$  in the flat torus of size  $L$ ,  $\mathbb{T}_L^\nu = \mathbb{R}^\nu / L \cdot \mathbb{Z}^\nu$ . (Just for simplicity, we will assume that the radii have the common value  $r$ .) The center of the  $i$ -th ball is denoted by  $q_i$  ( $\in \mathbb{T}_L^\nu$ ), its time derivative is  $v_i = \dot{q}_i$ ,  $i = 1, \dots, N$ . One uses the standard reduction of kinetic energy  $E = \frac{1}{2} \sum_{i=1}^N m_i \|v_i\|^2 = \frac{1}{2}$ . The arising configuration space (still without the removal of the scattering cylinders  $C_{i,j}$ ) is the torus

$$\mathbb{T}_L^{\nu N} = (\mathbb{T}_L^\nu)^N = \{(q_1, \dots, q_N) : q_i \in \mathbb{T}_L^\nu, i = 1, \dots, N\}$$

supplied with the Riemannian inner product (the so-called mass metric)

$$\langle v, v' \rangle = \sum_{i=1}^N m_i \langle v_i, v'_i \rangle \quad (2.2.1)$$

in its common tangent space  $\mathbb{R}^{\nu N} = (\mathbb{R}^\nu)^N$ . Now the Euklidean space  $\mathbb{R}^{\nu N}$  with the inner product (2.2.1) plays the role of  $\mathbb{R}^d$  in the original definition of cylindric billiards, see §2.1 above.

The generator subspace  $A_{i,j} \subset \mathbb{R}^{\nu N}$  ( $1 \leq i < j \leq N$ ) of the cylinder  $C_{i,j}$  (describing the collisions between the  $i$ -th and  $j$ -th balls) is given by the equation

$$A_{i,j} = \left\{ (q_1, \dots, q_N) \in (\mathbb{R}^\nu)^N : q_i = q_j \right\}, \quad (2.2.2)$$

see (4.3) in [S-Sz(2000)]. Its ortho-complement  $L_{i,j} \subset \mathbb{R}^{\nu N}$  is then defined by the equation

$$L_{i,j} = \left\{ (q_1, \dots, q_N) \in (\mathbb{R}^\nu)^N : q_k = 0 \text{ for } k \neq i, j, \text{ and } m_i q_i + m_j q_j = 0 \right\}, \quad (2.2.3)$$

see (4.4) in [S-Sz(2000)]. Easy calculation shows that the cylinder  $C_{i,j}$  (describing the overlap of the  $i$ -th and  $j$ -th balls) is indeed spherical and the radius of its base sphere is equal to  $r_{i,j} = 2r \sqrt{\frac{m_i m_j}{m_i + m_j}}$ , see §4, especially formula (4.6) in [S-Sz(2000)].

The structure lattice  $\mathcal{L} \subset \mathbb{R}^{\nu N}$  is clearly the lattice  $\mathcal{L} = (L \cdot \mathbb{Z}^\nu)^N = L \cdot \mathbb{Z}^{N\nu}$ .

Due to the presence of an extra invariant quantity  $I = \sum_{i=1}^N m_i v_i$ , one usually makes the reduction  $\sum_{i=1}^N m_i v_i = 0$  and, correspondingly, factorizes the configuration space with respect to uniform spatial translations:

$$(q_1, \dots, q_N) \sim (q_1 + a, \dots, q_N + a), \quad a \in \mathbb{T}_L^\nu. \quad (2.2.4)$$

The natural, common tangent space of this reduced configuration space is then

$$\mathcal{Z} = \left\{ (v_1, \dots, v_N) \in (\mathbb{R}^\nu)^N : \sum_{i=1}^N m_i v_i = 0 \right\} = \left( \bigcap_{i < j} A_{i,j} \right)^\perp = (\mathcal{A})^\perp \quad (2.2.5)$$

supplied again with the inner product (2.2.1), see also (4.1) and (4.2) in [Sz(2000)]. The base spaces  $L_{i,j}$  of (2.2.3) are obviously subspaces of  $\mathcal{Z}$ , and we take  $\tilde{A}_{i,j} = A_{i,j} \cap \mathcal{Z} = P_{\mathcal{Z}}(A_{i,j})$  as the ortho-complement of  $L_{i,j}$  in  $\mathcal{Z}$ . (Here  $P_{\mathcal{Z}}$  denotes the orthogonal projection onto the space  $\mathcal{Z}$ .)

Note that the configuration space of the reduced system (with the identification (2.2.4)) is naturally the torus  $\mathbb{R}^{\nu N} / (\mathcal{A} + L \cdot \mathbb{Z}^{\nu N}) = \mathcal{Z} / P_{\mathcal{Z}}(L \cdot \mathbb{Z}^{\nu N})$ .

### 2.3 Collision graphs

Let  $S^{[a,b]}x$  be a nonsingular, finite trajectory segment with the collisions  $\sigma_1, \dots, \sigma_n$  listed in time order. (Each  $\sigma_k$  is an unordered pair  $(i, j)$  of different labels  $i, j \in \{1, 2, \dots, N\}$ .) The graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  with vertex set  $\mathcal{V} = \{1, 2, \dots, N\}$  and set of edges  $\mathcal{E} = \{\sigma_1, \dots, \sigma_n\}$  is called the *collision graph* of the orbit segment  $S^{[a,b]}x$ . For a given positive number  $C$ , the collision graph  $\mathcal{G} = (\mathcal{V}, \mathcal{E})$  of the orbit segment  $S^{[a,b]}x$  will be called *C-rich* if  $\mathcal{G}$  contains at least  $C$  connected, consecutive (i.e., following one after the other in time, according to the time-ordering given by the trajectory segment  $S^{[a,b]}x$ ) subgraphs.

### 2.4 Trajectory branches

We are going to briefly describe the discontinuity of the flow  $\{S^t\}$  caused by a multiple collisions at time  $t_0$ . Assume first that the pre-collision velocities of the particles are given. What can we say about the possible post-collision velocities? Let us perturb the pre-collision phase point (at time  $t_0 - 0$ ) infinitesimally, so that the collisions at  $\sim t_0$  occur at infinitesimally different moments. By applying the collision laws to the arising finite sequence of collisions, we see that the post-collision velocities are fully determined by the time-ordering of the considered collisions. Therefore, the collection of all possible time-orderings of these collisions gives rise to a finite family of continuations of the trajectory beyond  $t_0$ . They are called the trajectory branches. It is quite clear that similar statements can be said regarding the evolution of a trajectory through a multiple collision in reverse time. Furthermore, it is also obvious that for any given phase point  $x_0 \in \mathbf{M}$  there are two,  $\omega$ -high trees  $\mathcal{T}_+$  and  $\mathcal{T}_-$  such that  $\mathcal{T}_+$  ( $\mathcal{T}_-$ ) describes all the possible



continuations of the positive (negative) trajectory  $S^{[0,\infty)}x_0$  ( $S^{(-\infty,0]}x_0$ ). (For the definitions of trees and for some of their applications to billiards, cf. the beginning of §5 in [K-S-Sz(1992)].) It is also clear that all possible continuations (branches) of the whole trajectory  $S^{(-\infty,\infty)}x_0$  can be uniquely described by all pairs  $(B_-, B_+)$  of  $\omega$ -high branches of the trees  $\mathcal{T}_-$  and  $\mathcal{T}_+$  ( $B_- \subset \mathcal{T}_-, B_+ \subset \mathcal{T}_+$ ).

Finally, we note that the trajectory of the phase point  $x_0$  has exactly two branches, provided that  $S^t x_0$  hits a singularity for a single value  $t = t_0$ , and the phase point  $S^{t_0} x_0$  does not lie on the intersection of more than one singularity manifolds. In this case we say that the trajectory of  $x_0$  has a “simple singularity”.

## 2.5 Neutral subspaces, advance, and sufficiency

Consider a nonsingular trajectory segment  $S^{[a,b]}x$ . Suppose that  $a$  and  $b$  are not moments of collision.

**Definition 2.5.1** *The neutral space  $\mathcal{N}_0(S^{[a,b]}x)$  of the trajectory segment  $S^{[a,b]}x$  at time zero ( $a < 0 < b$ ) is defined by the following formula:*

$$\mathcal{N}_0(S^{[a,b]}x) = \{W \in \mathcal{Z}: \exists(\delta > 0) \text{ such that } \forall \alpha \in (-\delta, \delta) \\ V(S^a(Q(x) + \alpha W, V(x))) = V(S^a x) \text{ and } V(S^b(Q(x) + \alpha W, V(x))) = V(S^b x)\}.$$

( $\mathcal{Z}$  is the common tangent space  $\mathcal{T}_q \mathbf{Q}$  of the parallelizable manifold  $\mathbf{Q}$  at any of its points  $q$ , while  $V(x)$  is the velocity component of the phase point  $x = (Q(x), V(x))$ .)

It is known (see (3) in §3 of [S-Ch (1987)]) that  $\mathcal{N}_0(S^{[a,b]}x)$  is a linear subspace of  $\mathcal{Z}$  indeed, and  $V(x) \in \mathcal{N}_0(S^{[a,b]}x)$ . The neutral space  $\mathcal{N}_t(S^{[a,b]}x)$  of the segment  $S^{[a,b]}x$  at time  $t \in [a, b]$  is defined as follows:

$$\mathcal{N}_t(S^{[a,b]}x) = \mathcal{N}_0(S^{[a-t, b-t]}(S^t x)).$$

It is clear that the neutral space  $\mathcal{N}_t(S^{[a,b]}x)$  can be canonically identified with  $\mathcal{N}_0(S^{[a,b]}x)$  by the usual identification of the tangent spaces of  $\mathbf{Q}$  along the trajectory  $S^{(-\infty, \infty)}x$  (see, for instance, §2 of [K-S-Sz(1990)]).

Our next definition is that of the advance. Consider a non-singular orbit segment  $S^{[a,b]}x$  with the symbolic collision sequence  $\Sigma = (\sigma_1, \dots, \sigma_n)$  ( $n \geq 1$ ), meaning that  $S^{[a,b]}x$  has exactly  $n$  collisions with  $\partial \mathbf{Q}$ , and the  $i$ -th collision ( $1 \leq i \leq n$ ) takes place at the boundary of the cylinder  $C_{\sigma_i}$ . For  $x = (Q, V) \in \mathbf{M}$  and  $W \in \mathcal{Z}$ ,  $\|W\|$  sufficiently small, denote  $T_W(Q, V) := (Q + W, V)$ .

**Definition 2.5.2** *For any  $1 \leq k \leq n$  and  $t \in [a, b]$ , the advance*

$$\alpha(\sigma_k): \mathcal{N}_t(S^{[a,b]}x) \rightarrow \mathbb{R}$$

*of the collision  $\sigma_k$  is the unique linear extension of the linear functional  $\alpha(\sigma_k)$  defined in a sufficiently small neighborhood of the origin of  $\mathcal{N}_t(S^{[a,b]}x)$  in the following way:*

$$\alpha(\sigma_k)(W) := t_k(x) - t_k(S^{-t} T_W S^t x).$$

Here  $t_k = t_k(x)$  is the time moment of the  $k$ -th collision  $\sigma_k$  on the trajectory of  $x$  after time  $t = a$ . The above formula and the notion of the advance functional

$$\alpha_k = \alpha(\sigma_k) : \mathcal{N}_t(S^{[a,b]}x) \rightarrow \mathbb{R}$$

has two important features:

- (i) If the spatial translation  $(Q, V) \mapsto (Q + W, V)$  is carried out at time  $t$ , then  $t_k$  changes linearly in  $W$ , and it takes place just  $\alpha_k(W)$  units of time earlier. (This is why it is called “advance”.)
- (ii) If the considered reference time  $t$  is somewhere between  $t_{k-1}$  and  $t_k$ , then the neutrality of  $W$  with respect to  $\sigma_k$  precisely means that

$$W - \alpha_k(W) \cdot V(x) \in A_{\sigma_k},$$

i.e., a neutral (with respect to the collision  $\sigma_k$ ) spatial translation  $W$  with the advance  $\alpha_k(W) = 0$  means that the vector  $W$  belongs to the generator space  $A_{\sigma_k}$  of the cylinder  $C_{\sigma_k}$ .

It is now time to bring up the basic notion of sufficiency (or, sometimes it is also called geometric hyperbolicity) of a trajectory (segment). This is the utmost important necessary condition for the proof of the fundamental theorem for algebraic semi-dispersive billiards, see Theorem 4.4 in [B-Ch-Sz-T(2002)].

### Definition 2.5.3

- (i) The nonsingular trajectory segment  $S^{[a,b]}x$  ( $a$  and  $b$  are supposed not to be moments of collision) is said to be sufficient if and only if the dimension of  $\mathcal{N}_t(S^{[a,b]}x)$  ( $t \in [a, b]$ ) is minimal, i.e.,  $\dim \mathcal{N}_t(S^{[a,b]}x) = 1$ .
- (ii) The trajectory segment  $S^{[a,b]}x$  containing exactly one singularity (a so-called “simple singularity”, see 2.4 above) is said to be sufficient if and only if both branches of this trajectory segment are sufficient.

**Definition 2.5.4** The phase point  $x \in \mathbf{M}$  with at most one (simple) singularity is said to be sufficient if and only if its whole trajectory  $S^{(-\infty, \infty)}x$  is sufficient, which means, by definition, that some of its bounded segments  $S^{[a,b]}x$  are sufficient.

In the case of an orbit  $S^{(-\infty, \infty)}x$  with a simple singularity, sufficiency means that both branches of  $S^{(-\infty, \infty)}x$  are sufficient.

## 2.6 No accumulation (of collisions) in finite time

By the results of Vaserstein [V(1979)], Galperin [G(1981)] and Burago-Ferleger-Kononenko [B-F-K(1998)], in a semi-dispersive billiard flow with the property (2.1.2) there can only be finitely many collisions in finite time intervals, see Theorem 1 in [B-F-K(1998)]. Thus, the dynamics is well defined as long as the trajectory does not hit more than one boundary components at the same time.

## 2.7 Slim sets

We are going to summarize the basic properties of codimension-two subsets  $A$  of a connected, smooth manifold  $M$  with a possible boundary. Since these subsets  $A$  are just those negligible in our dynamical discussions, we shall call them slim. As to a broader exposition of the issues, see [E(1978)] or §2 of [K-S-Sz(1991)].

Note that the dimension  $\dim A$  of a separable metric space  $A$  is one of the three classical notions of topological dimension: the covering (Čech-Lebesgue), the small inductive (Menger-Urysohn), or the large inductive (Brouwer-Čech) dimension. As it is known from general topology, all of them are the same for separable metric spaces.

**Definition 2.7.1** A subset  $A$  of  $M$  is called slim if and only if  $A$  can be covered by a countable family of codimension-two (i.e., at least two) closed sets of  $\mu$ -measure zero, where  $\mu$  is a smooth measure on  $M$ . (Cf. Definition 2.12 of [K-S-Sz(1991)].)

**Property 2.7.2** The collection of all slim subsets of  $M$  is a  $\sigma$ -ideal, that is, countable unions of slim sets and arbitrary subsets of slim sets are also slim.

**Proposition 2.7.3. (Locality)** A subset  $A \subset M$  is slim if and only if for every  $x \in A$  there exists an open neighborhood  $U$  of  $x$  in  $M$  such that  $U \cap A$  is slim. (Cf. Lemma 2.14 of [K-S-Sz(1991)].)

**Property 2.7.4** A closed subset  $A \subset M$  is slim if and only if  $\mu(A) = 0$  and  $\dim A \leq \dim M - 2$ .

**Property 2.7.5. (Integrability)** If  $A \subset M_1 \times M_2$  is a closed subset of the product of two smooth manifolds with possible boundaries, and for every  $x \in M_1$  the set

$$A_x = \{y \in M_2 : (x, y) \in A\}$$

is slim in  $M_2$ , then  $A$  is slim in  $M_1 \times M_2$ .

The following propositions characterize the codimension-one and codimension-two sets.

**Proposition 2.7.6** For any closed subset  $S \subset M$  the following three conditions are equivalent:

- (i)  $\dim S \leq \dim M - 2$ ;
- (ii)  $\text{int} S = \emptyset$  and for every open connected set  $G \subset M$  the difference set  $G \setminus S$  is also connected;
- (iii)  $\text{int} S = \emptyset$  and for every point  $x \in M$  and for any open neighborhood  $V$  of  $x$  in  $M$  there exists a smaller open neighborhood  $W \subset V$  of the point  $x$  such that for every pair of points  $y, z \in W \setminus S$  there is a continuous curve  $\gamma$  in the set  $V \setminus S$  connecting the points  $y$  and  $z$ .

(See Theorem 1.8.13 and Problem 1.8.E of [E(1978)].)

**Proposition 2.7.7** For any subset  $S \subset M$  the condition  $\dim S \leq \dim M - 1$  is equivalent to  $\text{int} S = \emptyset$ . (See Theorem 1.8.10 of [E(1978)].)

We recall an elementary, but important lemma (Lemma 4.15 of [K-S-Sz(1991)]). Let  $R_2$  be the set of phase points  $x \in \mathbf{M} \setminus \partial\mathbf{M}$  such that the trajectory  $S^{(-\infty, \infty)}x$  has more than one singularities.

**Proposition 2.7.8** The set  $R_2$  is a countable union of codimension-two smooth sub-manifolds of  $M$  and, being such, it is slim.

The next lemma establishes the most important property of slim sets which gives us the fundamental geometric tool to connect the open ergodic components of billiard flows.

**Proposition 2.7.9** If  $M$  is connected, then the complement  $M \setminus A$  of a slim  $F_\sigma$  set  $A \subset M$  is an arc-wise connected ( $G_\delta$ ) set of full measure. (See Property 3 of §4.1 in [K-S-Sz(1989)]. The  $F_\sigma$  sets are, by definition, the countable unions of closed sets, while the  $G_\delta$  sets are the countable intersections of open sets.)

## 2.8 The subsets $\mathbf{M}^0$ and $\mathbf{M}^\#$

Denote by  $\mathbf{M}^\#$  the set of all phase points  $x \in \mathbf{M}$  for which the trajectory of  $x$  encounters infinitely many non-tangential collisions in both time directions. The trajectories of the points  $x \in \mathbf{M} \setminus \mathbf{M}^\#$  are lines: the motion is linear and uniform, see the appendix of [Sz(1994)]. It is proven in lemmas A.2.1 and A.2.2 of [Sz(1994)] that the closed set  $\mathbf{M} \setminus \mathbf{M}^\#$  is a finite union of hyperplanes. It is also proven in [Sz(1994)] that, locally, the two sides of a hyper-planar component of  $\mathbf{M} \setminus \mathbf{M}^\#$  can be connected by a positively measured beam of trajectories, hence, from the point of view of ergodicity, in this paper it is enough to show that the connected components of  $\mathbf{M}^\#$  entirely belong to one ergodic component. This is what we are going to do in this paper.

Denote by  $\mathbf{M}^0$  the set of all phase points  $x \in \mathbf{M}^\#$  the trajectory of which does not hit any singularity, and use the notation  $\mathbf{M}^1$  for the set of all phase points  $x \in \mathbf{M}^\#$  whose orbit contains exactly one, simple singularity. According to Proposition 2.7.8, the set  $\mathbf{M}^\# \setminus (\mathbf{M}^0 \cup \mathbf{M}^1)$  is a countable union of smooth, codimension-two ( $\geq 2$ ) submanifolds of  $\mathbf{M}$ , and, therefore, this set may be discarded in our study of ergodicity, please see also the properties of slim sets above. Thus, we will restrict our attention to the phase points  $x \in \mathbf{M}^0 \cup \mathbf{M}^1$ .

## 2.9 The “Chernov-Sinai Ansatz”

An essential precondition for the theorem on local ergodicity by Bálint-Chernov-Szász-Tóth (Theorem 4.4 of [B-Ch-Sz-T(2002)]) is the so-called “Chernov-Sinai Ansatz” which we are going to formulate below. Denote by  $\mathcal{SR}^+ \subset \partial\mathbf{M}$  the set of all phase points  $x_0 = (q_0, v_0) \in \partial\mathbf{M}$  corresponding to singular reflections (a

tangential or a double collision at time zero) supplied with the post-collision (outgoing) velocity  $v_0$ . It is well known that  $\mathcal{SR}^+$  is a compact cell complex with dimension  $2d - 3 = \dim \mathbf{M} - 2$ . It is also known (see Lemma 4.1 in [K-S-Sz(1990)]) that for  $\nu$ -almost every phase point  $x_0 \in \mathcal{SR}^+$  the forward orbit  $S^{(0,\infty)}x_0$  does not hit any further singularity. (Here  $\nu$  is the Riemannian volume of  $\mathcal{SR}^+$  induced by the restriction of the natural Riemannian metric of  $\mathbf{M}$ .) The Chernov-Sinai Ansatz postulates that for  $\nu$ -almost every  $x_0 \in \mathcal{SR}^+$  the forward orbit  $S^{(0,\infty)}x_0$  is sufficient (geometrically hyperbolic).

## 2.10 The theorem on local ergodicity

The theorem on local ergodicity by Bálint-Chernov-Szász-Tóth (Theorem 4.4 of [B-Ch-Sz-T(2002)]) claims the following: Let  $(\mathbf{M}, \{S^t\}_{t \in \mathbb{R}}, \mu)$  be a semi-dispersive billiard flow with (2.1.1)–(2.1.2) and with the property that the smooth components of the boundary  $\partial \mathbf{Q}$  of the configuration space are algebraic hyper-surfaces. (The cylindric billiards automatically fulfill this algebraicity condition.) Assume – further – that the Chernov-Sinai Ansatz holds true, and a phase point  $x_0 \in (\mathbf{M} \setminus \partial \mathbf{M}) \cap \mathbf{M}^\#$  is given with the properties

(i)  $S^{(-\infty,\infty)}x$  has at most one singularity,

and

(ii)  $S^{(-\infty,\infty)}x$  is sufficient.

Then some open neighborhood  $U_0 \subset \mathbf{M}$  of  $x_0$  belongs to a single ergodic component of the flow  $(\mathbf{M}, \{S^t\}_{t \in \mathbb{R}}, \mu)$ . (Modulo the zero sets, of course.)

## 3 Non-sufficiency occurs on a codimension-two set. The case $\nu \geq 3$

The opening part of this section contains a slightly modified version of Lemma 4.43 from [S-Sz(1999)]. The reason why we had to modify the recursion for the sequence  $C(N)$  (from  $C(N) = (N/2) \cdot \max\{C(N-1), 3\}$  to  $C(N) = (N/2) \cdot (2C(N-1) + 1)$ ) is that our Corollary 3.5 (below) requires  $(2C(N) + 1)$ -richness instead of the usual  $C(N)$ -richness. In the present paper the sequence  $C(N)$  always denotes the one defined by the recursion in Lemma 3.1 instead of the one defined in Lemma 4.43 of [S-Sz(1999)]. This should not cause any confusion.

We note that the upcoming lemma is purely combinatorial.

**Lemma 3.1** Define the sequence of positive numbers  $C(N)$  recursively by taking  $C(2) = 1$  and  $C(N) = (N/2) \cdot (2C(N-1) + 1)$  for  $N \geq 3$ . Let  $N \geq 3$ , and suppose that the symbolic collision sequence  $\Sigma = (\sigma_1, \dots, \sigma_n)$  for  $N$  particles is  $C(N)$ -rich. Then we can find a particle, say the one with label  $N$ , and two indices  $1 \leq p < q \leq n$  such that

- (i)  $N \in \sigma_p \cap \sigma_q$ ,
- (ii)  $N \notin \bigcup_{j=p+1}^{q-1} \sigma_j$ ,
- (iii)  $\sigma_p = \sigma_q \implies (\exists j) (p < j < q \text{ \& } \sigma_p \cap \sigma_j \neq \emptyset)$ , and
- (iv)  $\Sigma'$  is  $(2C(N-1)+1)$ -rich on the vertex set  $\{1, \dots, N-1\}$ .

(Here, just as in the case of derived schemes, we denote by  $\Sigma'$  the symbolic sequence that can be obtained from  $\Sigma$  by discarding all edges containing  $N$ .)

*Proof.* The hypothesis on  $\Sigma$  implies that there exist subsequences  $\Sigma_1, \dots, \Sigma_r$  of  $\Sigma$  with the following properties:

- (1) For  $1 \leq i < j \leq r$  every collision of  $\Sigma_i$  precedes every collision of  $\Sigma_j$ ,
- (2) the graph of  $\Sigma_i$  ( $1 \leq i \leq r$ ) is a tree (a connected graph without loop) on the vertex set  $\{1, \dots, N\}$ , and
- (3)  $r \geq C(N)$ .

Since every tree contains at least two vertices with degree one and  $C(N) = (N/2) \cdot \{2C(N-1)+1\}$ , there is a vertex, say the one labeled by  $N$ , such that  $N$  is a degree-one vertex of  $\Sigma_{i(1)}, \dots, \Sigma_{i(t)}$ , where  $1 \leq i(1) < \dots < i(t) \leq r$  and  $t \geq 2C(N-1)+1$ . Thus (iv) obviously holds.

Let  $\sigma_{p'}$  the edge of  $\Sigma_{i(1)}$  that contains  $N$  and, similarly, let  $\sigma_{q'}$  be the edge of  $\Sigma_{i(t)}$  containing the vertex  $N$ . Then the fact  $t \geq 3$  ensures that the following properties hold:

- (i)'  $N \in \sigma_{p'} \cap \sigma_{q'}$ ,
- (iii)'  $\sigma_{p'} = \sigma_{q'} \implies \exists j p' < j < q' \text{ \& } \sigma_{p'} \cap \sigma_j \neq \emptyset, \sigma_j \neq \sigma_{p'}$ .

Let  $\sigma_p, \sigma_q$  ( $1 \leq p < q \leq n$ ) be a pair of edges  $\sigma_{p'}, \sigma_{q'}$  ( $1 \leq p' < q' \leq n$ ) fulfilling (i)' and (iii)' and having the minimum possible value of  $q' - p'$ . Elementary inspection shows that then (ii) must also hold for  $\sigma_p, \sigma_q$ . Lemma 3.5.1 is now proved.  $\square$

Let us fix a triplet  $(\Sigma, \mathcal{A}, \bar{\tau})$  of the discrete (combinatorial) orbit structure with Property (A) (just as in [S-Sz(1999)], see Definition 3.31 there), and assume that  $\Sigma = (\sigma_1, \dots, \sigma_n)$  is  $C(N)$ -rich, i.e., it contains at least  $C(N)$  consecutive, connected collision graphs. We also consider the complex analytic manifold  $\Omega(\Sigma, \mathcal{A}, \bar{\tau})$  of all complex  $(\Sigma, \mathcal{A}, \bar{\tau})$ -orbits  $\omega$  (Definition 3.20 in [S-Sz(1999)]) and the open, dense, connected domain  $D(\Sigma, \mathcal{A}, \bar{\tau}) \subset \mathbb{C}^{(2\nu+1)N+1}$  of all allowable initial data  $\vec{x} = \vec{x}(\omega)$ , see Definition 3.18 in [S-Sz(1999)]. Let, finally,  $Q(\vec{x})$  be a common *irreducible* divisor of the polynomials  $P_1(\vec{x}), \dots, P_s(\vec{x})$  from (4.3) in [S-Sz(1999)]. (If such a common divisor exists.) In this section we will need several results about such common irreducible divisors  $Q(\vec{x})$  of the polynomials  $P_1(\vec{x}), \dots, P_s(\vec{x})$ .

The first of them, as it is classically known from algebraic geometry (see, for example, [M(1976)]), is that the solution set  $V = \{Q(\vec{x}) = 0\}$  of the equation  $Q(\vec{x}) = 0$  is a so-called irreducible (or, indecomposable) complex algebraic variety of codimension 1 in  $\mathbb{C}^{(2\nu+1)N+1}$ , which means that  $V$  is not the union of two, proper algebraic sub-varieties. Secondly, the smooth part  $S$  of  $V$  turns out to be a connected complex analytic manifold, while the non-smooth part  $V \setminus S$  of  $V$

is a complex algebraic variety of dimension strictly less than  $\dim V = (2\nu + 1)N$ , see [M(1976)]. Finally, if a polynomial  $R(\vec{x})$  vanishes on  $V$ , then  $Q(\vec{x})$  must be a divisor of  $R(\vec{x})$ . (The last statement is a direct consequence of Hilbert's theorem on Zeroes, see again [M(1976)].)

The first result, specific to our current dynamical situation, is

**Proposition 3.2** The polynomials  $P_1(\vec{x}), \dots, P_s(\vec{x})$  of (4.3) in [S-Sz(1999)] are homogeneous in the masses  $m_1, \dots, m_N$  and, consequently, any common divisor  $Q(\vec{x})$  of these polynomials is also homogeneous in the masses.

*Proof.* It is clear that the complex dynamics encoded in the orbits  $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$  only depends on the ratios of masses  $m_2/m_1, m_3/m_1, \dots, m_N/m_1$ . Consequently, all algebraic functions  $f_i(\vec{x})$  ( $i = 1, \dots, s$ ) featuring the proof of Lemma 4.2 of [S-Sz(1999)] are homogeneous of degree 0 in the masses. Since the rational function

$$\frac{P_i(\vec{x})}{Q_i(\vec{x})} \in \mathbb{K}_0 = \mathbb{C}(\vec{x})$$

is the product  $\alpha$  of all conjugates of  $f_i(\vec{x})$  (see the proof of the lemma just cited), we get that  $\frac{P_i(\vec{x})}{Q_i(\vec{x})}$  is also homogeneous of degree 0 in the variables  $m_1, \dots, m_N$ . Then elementary algebra yields that both  $P_i(\vec{x})$  and  $Q_i(\vec{x})$  are homogeneous (of the same degree) in the masses. Since any factor of a homogeneous polynomial is easily seen to be also homogeneous, we get that the common divisor  $Q(\vec{x})$  of  $P_1(\vec{x}), \dots, P_s(\vec{x})$  is also homogeneous in the variables  $m_1, \dots, m_N$ .  $\square$

Our next result, specific to our dynamics, that will be needed later is

**Proposition 3.3** Let  $\nu \geq 3$ ,  $(\Sigma, \mathcal{A}, \vec{\tau})$  be a discrete orbit structure with Property (A) and a  $C(N)$ -rich symbolic collision sequence  $\Sigma = (\sigma_1, \dots, \sigma_n)$ . Denote by  $P_1(\vec{x}), \dots, P_s(\vec{x})$  the polynomials of (4.3) in [S-Sz(1999)] just as before, and let  $Q(\vec{x})$  be a common irreducible divisor of  $P_1(\vec{x}), \dots, P_s(\vec{x})$ . (If such a common divisor exists.) Let, finally,  $(\overline{\Sigma}, \overline{\mathcal{A}}, \vec{\rho})$  be an extended discrete orbit structure with Property (A) and an extended collision sequence  $\overline{\Sigma} = (\sigma_0, \sigma_1, \dots, \sigma_n)$ . We claim that the irreducible (indecomposable) solution set  $V$  of the equation  $Q(\vec{x}) = 0$  cannot even locally coincide with any of the following singularity manifolds  $C$  defined by one of the following equations:

- (1)  $\|v_{i_0}^0 - v_{j_0}^0\|^2 = 0$ ,
- (2)  $\langle v_{i_0}^0 - v_{j_0}^0; \tilde{q}_{i_0}^0 - \tilde{q}_{j_0}^0 - L \cdot a_0 \rangle = 0$ ,
- (3)  $m_{i_0} + m_{j_0} = 0$ ,

i.e., the irreducible polynomial  $Q(\vec{x})$  is not equal to any of the (irreducible) polynomials on the left-hand sides of (1)–(3). (In [S-Sz(1999)] these equations feature Definition 3.18 of the domain  $D(\overline{\Sigma}, \overline{\mathcal{A}}, \vec{\rho})$ .) Consequently, the open subset  $V \cap D(\overline{\Sigma}, \overline{\mathcal{A}}, \vec{\rho})$  of  $V$  is connected and dense in  $V$ .

**Remark.** The first point where we (implicitly) use the condition  $\nu \geq 3$  is the irreducibility of the polynomial  $\|v_{i_0}^0 - v_{j_0}^0\|^2$  on the left-hand side of (1). Indeed, in the case  $\nu = 2$  this polynomial splits as

$$\begin{aligned} \|v_{i_0}^0 - v_{j_0}^0\|^2 &\equiv \left[ (v_{i_0}^0)_1 - (v_{j_0}^0)_1 + \sqrt{-1} \left( (v_{i_0}^0)_2 - (v_{j_0}^0)_2 \right) \right] \\ &\quad \cdot \left[ (v_{i_0}^0)_1 - (v_{j_0}^0)_1 - \sqrt{-1} \left( (v_{i_0}^0)_2 - (v_{j_0}^0)_2 \right) \right]. \end{aligned}$$

However, it is easy to see that, in the case  $\nu \geq 3$ , all polynomials on the left of (1)–(3) are indeed irreducible.

*Proof.* First of all, we slightly reformulate the negation of the statement of the proposition.

Fix one of the three equations of (1)–(3) above, and denote the irreducible polynomial on its left-hand side by  $R(\vec{x})$ . By using the quadratic (or linear) equation  $R(\vec{x}) = 0$ , we eliminate one variable  $x_j$  out of  $\vec{x}$  by expressing it as an algebraic function  $x_j = g(\vec{y})$  of the remaining variables  $\vec{y}$  of  $\vec{x}$ , so that the algebraic function  $g$  only contains finitely many field operations and (at most one) square root. After this elimination  $x_j = g(\vec{y})$ , the meaning of  $R(\vec{x}) \equiv Q(\vec{x})$  (i.e., the negation of the assertion of the proposition) is that all algebraic functions  $f_i(\vec{x}) \equiv \tilde{f}_i(\vec{y})$  in the proof of Lemma 4.2 of [S-Sz(1999)] ( $i = 1, \dots, s$ , constructed for  $(\Sigma, \mathcal{A}, \vec{\tau})$ , not for  $(\bar{\Sigma}, \bar{\mathcal{A}}, \bar{\rho})$ ) are identically zero in terms of  $\vec{y}$ , meaning that every complex orbit segment  $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$ , with the initial data  $\vec{x}(\omega)$  in the solution set of  $R(\vec{x}) = 0$ , is non-sufficient, see also the “Dichotomy Corollary” 4.7 in [S-Sz(1999)]. Thus, the negation of the proposition means that no orbit segment  $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$  in the considered singularity is sufficient.

Now we carry out an induction on the number of balls  $N$  quite in the spirit of the proof of Key Lemma 4.1 of [S-Sz(1999)]. Indeed, the statement of the proposition is obviously true in the case  $N = 2$ , for in that case there are no non-sufficient (complex) trajectories  $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$ , i.e., the greatest common divisor of the polynomials  $P_1(\vec{x}), \dots, P_s(\vec{x})$  is 1.

Assume now that  $N \geq 3$ ,  $\nu \geq 3$ , and the statement of Proposition 3.3 has been proven for all values  $N' < N$ . Suppose, however, that the statement is false for some  $(\Sigma, \mathcal{A}, \vec{\tau})$  and extension  $(\bar{\Sigma}, \bar{\mathcal{A}}, \bar{\rho})$  with  $N$  balls and Property (A), i.e., that there exists a common irreducible divisor  $Q(\vec{x})$  of all the polynomials  $P_1(\vec{x}), \dots, P_s(\vec{x})$ , and  $Q(\vec{x})$  happens to be one of the irreducible polynomials on the left-hand side of (1), (2), or (3). By using the  $C(N)$ -richness of  $\Sigma = (\sigma_1, \dots, \sigma_n)$ , we select a suitable label  $k_0 \in \{1, 2, \dots, N\}$ , say  $k_0 = N$ , for the substitution  $m_N = 0$  along the lines of Lemma 3.1 above, by also ensuring the existence of the derived schemes  $(\Sigma', \mathcal{A}', \vec{\tau}')$  and  $(\bar{\Sigma}', \bar{\mathcal{A}}', \bar{\rho}')$  for the  $(N-1)$ -ball-system  $\{1, 2, \dots, N-1\}$  and properties (i)–(iv) (of Lemma 3.1) for  $\Sigma'$ , see Corollary 4.35 of [S-Sz(1999)] and Lemma 3.1 above. Denote by  $\tilde{Q}(\vec{x})$  the polynomial obtained from  $Q(\vec{x})$  after the substitution  $m_N = 0$ .

**Lemma 3.4** The polynomial  $\tilde{Q}(\vec{x})$  is not constant.



*Proof.* Assume that  $\tilde{Q}(\vec{x}) \equiv c \in \mathbb{C}$ . The case  $c = 0$  means that  $m_N$  is a divisor of  $Q(\vec{x})$ , thus  $m_N \equiv Q(\vec{x})$  (for  $Q(\vec{x})$  is irreducible), which is impossible, since  $Q(\vec{x})$  has to be one of the polynomials on the left-hand side of (1), (2), or (3).

If, however,  $c \neq 0$ , then we have that  $Q(\vec{x}) \equiv c + m_N S(\vec{x})$  with some nonzero polynomial  $S(\vec{x})$ . ( $S(\vec{x})$  has to be non-zero, otherwise  $Q(\vec{x})$  would be a constant, not an irreducible polynomial.) However, this contradicts to the proved homogeneity of the polynomial  $Q(\vec{x})$  in the masses, see Proposition 3.2 above. This finishes the proof of the lemma.  $\square$

**Remark.** If one takes a look at the equations (1), (2), (3), he/she immediately realizes that either  $\tilde{Q}(\vec{x}) \equiv Q(\vec{x})$  (when  $Q(\vec{x})$  is the polynomial on the left-hand side of (1) or (2), or  $Q(\vec{x}) \equiv m_{i_0} + m_{j_0}$  and  $N \notin \{i_0, j_0\}$ ), or  $\tilde{Q}(\vec{x}) \equiv m_{i_0}$  when  $Q(\vec{x}) \equiv m_{i_0} + m_{j_0}$  and  $N = j_0$ . In this way one can directly and easily verify Lemma 3.4 without the above “involved” algebraic proof. The reason why we still included the above proof is that later on in this section (in the proof of Sub-lemma 3.7) we will need the idea of the presented proof.

The next lemma will use

**Definition 3.5** Suppose that two indices  $1 \leq p < q \leq n$  and two labels of balls  $i, j \in \{1, \dots, N\}$  are given with the additional requirement that if  $i = j$ , then  $i \in \bigcup_{l=p+1}^{q-1} \sigma_l$ . Following the proof of Lemma 4.2 of [S-Sz(1999)], denote by  $Q_1(\vec{x}), Q_2(\vec{x}), \dots, Q_\nu(\vec{x})$  ( $\vec{x} \in \mathbb{C}^{(2\nu+1)N+1}$ ) the polynomials with the property that for every vector of initial data  $\vec{x} \in D(\Sigma, \mathcal{A}, \vec{\tau})$  and for every  $k, k = 1, \dots, \nu$ , the following equivalence holds true:

$$\begin{aligned} \left( \exists \omega \in \Omega \text{ such that } \vec{x}(\omega) = \vec{x} \text{ \& } (v_i^p(\omega))_k = \left( v_j^{q-1}(\omega) \right)_k \right) \\ \iff Q_k(\vec{x}) = 0. \end{aligned}$$

Our next lemma is a strengthened version of Lemma 4.39 of [S-Sz(1999)]:

**Lemma 3.6** Assume that the combinatorial-algebraic scheme  $(\Sigma, \mathcal{A}, \vec{\tau})$  has Property (A), and use the assumptions and notations of the above definition.

We claim that the polynomials  $Q_1(\vec{x}), Q_2(\vec{x}), \dots, Q_\nu(\vec{x})$  do not have any non-constant common divisor. In other words, the equality  $v_i^p(\omega) = v_j^{q-1}(\omega)$  only takes place on an algebraic variety with at least two codimensions.

**Remark.** Lemma 4.39 of [S-Sz(1999)] asserted that at least one of the above polynomials  $Q_k(\vec{x})$  is nonzero. Then, by the permutation symmetry of the components  $k \in \{1, \dots, \nu\}$ , all of these polynomials are actually nonzero.

*Proof.* Induction on the number  $N \geq 2$ .

1. Base of the induction,  $N = 2$ : First of all, by performing the substitution  $L = 0$ , we can annihilate all adjustment vectors, see (I), (IV), (VII) of Lemma 4.21 in [S-Sz(1999)], and Remark 4.22 there. Then, an elementary inspection shows that for any selection of *positive real* masses  $(m_1, m_2)$ , indeed, the

equality  $v_i^p(\omega) = v_j^{q-1}(\omega)$  only occurs on a manifold with  $\nu - 1$  ( $\geq 2$ ) codimensions in the section  $\Omega(\Sigma, \mathcal{A}, \vec{\tau}, \vec{m})$  of  $\Omega(\Sigma, \mathcal{A}, \vec{\tau})$  corresponding to the selected masses, since any trajectory segment of a two-particle system with positive masses and  $\mathcal{A} = 0$  has a very nice, totally real (and essentially  $\nu$ -dimensional) representation in the relative coordinates of the particles: The consecutive, elastic bounces of a point particle moving uniformly inside a ball of radius  $2r$  of  $\mathbb{R}^\nu$ . Therefore, the statement of the lemma is true for  $N = 2$ .

Assume now that  $N \geq 3$ , and the lemma has been successfully proven for all smaller numbers of balls. By re-labeling the particles, if necessary, we can achieve that

- (i)  $N \neq i, N \neq j$  and
- (ii) if  $i = j$ , then the ball  $i$  has at least one collision between  $\sigma_p$  and  $\sigma_q$  with a particle different from  $N$ .

For the fixed combinatorial scheme  $(\Sigma, \mathcal{A}, \vec{\tau})$ , select a derived scheme  $(\Sigma', \mathcal{A}', \vec{\tau}')$  corresponding to the substitution  $m_N = 0$ , see Definition 4.11 and Corollary 4.35 in [S-Sz(1999)].

Our induction step is going to be a proof by contradiction. Assume, therefore, that the nonzero polynomials  $Q_1(\vec{x}), Q_2(\vec{x}), \dots, Q_\nu(\vec{x})$  do have a common irreducible divisor  $R(\vec{x})$ . According to Proposition 3.2 above, the (irreducible) polynomial  $R(\vec{x})$  is homogeneous in the variables  $m_1, \dots, m_N$ . Denote by  $\tilde{R}(\vec{x})$  the polynomial that we obtain from  $R(\vec{x})$  after the substitution  $m_N = 0$ . Similarly to Lemma 3.4 above, we claim

**Sub-lemma 3.7.** The polynomial  $\tilde{R}(\vec{x})$  is not constant.

**Remark.** The reason why we cannot simply apply Lemma 3.4 is that in the proof of that lemma we used the assumption that the irreducible polynomial  $Q(\vec{x})$  was one of the polynomials on the left-hand side of (1), (2), or (3) of Proposition 3.3. Right here we do not have such an assumption.

*Proof.* Suppose that  $\tilde{R}(\vec{x}) \equiv c$ , where  $c \in \mathbb{C}$  is a constant, i.e.,  $R(\vec{x}) \equiv c + m_N \cdot S(\vec{x})$ . In the case  $c = 0$  the polynomial  $m_N \equiv R(\vec{x})$  would be a common divisor of all the polynomials  $Q_1(\vec{x}), Q_2(\vec{x}), \dots, Q_\nu(\vec{x})$ , meaning that in the considered  $N$ -ball system  $\Omega(\Sigma, \mathcal{A}, \vec{\tau})$  the equation  $m_N(\omega) = 0$  implies the equality  $v_i^p(\omega) = v_j^{q-1}(\omega)$ . This, in turn, means that in the  $(N - 1)$ -ball system  $\{1, \dots, N - 1\}$  (with the discrete algebraic scheme  $(\Sigma', \mathcal{A}', \vec{\tau}')$ ) the equality  $v_i^p(\omega) = v_j^{q-1}(\omega)$  is an identity, thus contradicting to the induction hypothesis.

Therefore  $c \neq 0$ , and in the expansion  $R(\vec{x}) \equiv c + m_N \cdot S(\vec{x})$  of the irreducible polynomial  $R(\vec{x})$  we certainly have  $S(\vec{x}) \not\equiv 0$ , and this means that  $R(\vec{x})$  is not homogeneous in the mass variables, thus contradicting to Proposition 3.2. This finishes the proof of the sub-lemma.  $\square$

**Finishing the proof of Lemma 3.6.** Denote by  $\tilde{Q}_k(\vec{x})$  the polynomial that we obtain from  $Q_k(\vec{x})$  after the substitution  $m_N = 0$  ( $k = 1, \dots, \nu$ ), and by  $T_k(\vec{x})$  the polynomial constructed for the  $(N - 1)$ -ball system  $\{1, \dots, N - 1\}$  (with the discrete algebraic scheme  $(\Sigma', \mathcal{A}', \vec{\tau}')$ ) along the lines of Lemma 4.2 of [S-Sz(1999)],

describing the event  $(v_i^p(\omega))_k = (v_j^{q-1}(\omega))_k$  in this subsystem ( $k = 1, \dots, \nu$ ). It follows from the induction hypothesis that the zero set  $W_k$  of the polynomial  $\tilde{Q}_k(\vec{x})$  (in the phase space of the  $(N-1)$ -ball system  $(\Sigma', \mathcal{A}', \vec{\tau}')$ ) has a codimension-two intersection with the singularities of the  $(\Sigma', \mathcal{A}', \vec{\tau}')$  system. Indeed, otherwise we would have  $(v_i^p)_k \equiv (v_j^{q-1})_k$  on some (irreducible) singularity manifold of the  $(\Sigma', \mathcal{A}', \vec{\tau}')$  subsystem. Then, by the symmetry with respect to the coordinates  $k = 1, 2, \dots, \nu$ , we would have  $v_i^p \equiv v_j^{q-1}$  on a codimension-one singularity of the subsystem  $(\Sigma', \mathcal{A}', \vec{\tau}')$ , contradicting to the induction hypothesis. This means that the polynomial  $T_k(\vec{x})$  vanishes on the zero set  $W_k$  of  $\tilde{Q}_k(\vec{x})$ , so the non-constant common divisor  $\tilde{R}(\vec{x})$  of  $\tilde{Q}_k(\vec{x})$  is a common divisor of  $T_1(\vec{x}), \dots, T_\nu(\vec{x})$ , contradicting to the induction hypothesis. This finishes the proof of Lemma 3.6.  $\square$

**Continuing the proof of Proposition 3.3.** Denote by  $\tilde{P}_1(\vec{x}), \dots, \tilde{P}_t(\vec{x})$  the “ $P_i$  polynomials” of the  $N$ -ball system  $(\Sigma, \mathcal{A}, \vec{\tau})$  with the constraint  $m_N = 0$  constructed the same way as the polynomials  $P_1(\vec{x}), \dots, P_s(\vec{x})$  in (4.3) of [S-Sz(1999)] for the general case  $m_N \in \mathbb{C}$ , see also the proof of Lemma 4.2 in the cited paper. It follows from the algebraic construction of these polynomials that the irreducible polynomial  $\tilde{Q}(\vec{x})$  is a common divisor of  $\tilde{P}_1(\vec{x}), \dots, \tilde{P}_t(\vec{x})$ . Recall that, according to our indirect assumption made right before Lemma 3.4,  $Q(\vec{x})$  is a common, irreducible divisor of the polynomials  $P_1(\vec{x}), \dots, P_s(\vec{x})$  and, at the same time,  $Q(\vec{x})$  is one of the polynomials on the left-hand side of (1), (2), or (3) in Proposition 3.3. The polynomial  $\tilde{Q}(\vec{x})$  was obtained from  $Q(\vec{x})$  by the substitution  $m_N = 0$ .

Let us focus now on Lemma 4.9 of [S-Sz(1999)]. The non-sufficiency of the  $N$ -ball system  $(\Sigma, \mathcal{A}, \vec{\tau})$  with the side condition  $m_N = 0$  comes from two sources: Either from the parallelity of the relative velocities in (2) of Lemma 4.9, or from the non-sufficiency of the  $(N-1)$ -ball part of the orbit segment  $\omega$  with the combinatorial scheme  $(\Sigma', \mathcal{A}', \vec{\tau}')$ . The first case takes place on a complex algebraic set of (at least) 2 codimensions, thanks to our original assumption  $\nu \geq 3$  and Lemma 3.6 above. Concerning the application of the “non-equality” Lemma 3.6 above, we note here that once the velocities  $v_{i_p}^p(\omega)$  and  $v_{i_q}^{q-1}(\omega)$  are not equal, the relative velocities  $v_N^p(\omega) - v_{i_p}^p(\omega)$  and  $v_N^{q-1}(\omega) - v_{i_q}^{q-1}(\omega) = v_N^p(\omega) - v_{i_q}^{q-1}(\omega)$  are not parallel, unless the common velocity  $v_N^p(\omega) = v_N^{q-1}(\omega)$  belongs to the complex line connecting the different velocities  $v_{i_p}^p(\omega)$  and  $v_{i_q}^{q-1}(\omega)$ , which is a codimension- $(\nu-1)$  condition on the velocity  $v_N^p(\omega)$ . Therefore, the equation  $\tilde{Q}(\vec{x}) = 0$  with the irreducible common divisor  $\tilde{Q}(\vec{x})$  of the polynomials  $\tilde{P}_1(\vec{x}), \dots, \tilde{P}_t(\vec{x})$  can only describe the non-sufficiency of the  $(\Sigma', \mathcal{A}', \vec{\tau}')$ -part of the system, thus  $\tilde{Q}(\vec{x})$  should lack the kinetic and mass variables corresponding to the ball with label  $N$ , as the following sub-lemma states:

**Sub-lemma 3.8.** The irreducible common divisor  $\tilde{Q}(\vec{x})$  of the polynomials

$$\tilde{P}_1(\vec{x}), \dots, \tilde{P}_t(\vec{x})$$

does not contain the variables with label  $N$ .

*Proof.* Let  $D = D(\Sigma, \mathcal{A}, \vec{\tau} | m_N = 0) \subset \mathbb{C}^{(2\nu+1)N}$  be the open, connected and dense domain in  $\mathbb{C}^{(2\nu+1)N}$  defined analogously to Definition 3.18 (of [S-Sz(1999)]) but incorporating the constraint  $m_N = 0$ , see also Lemma 3.19 in [S-Sz(1999)]. Let, further,  $\mathcal{N} \subset D$  be a small, complex analytic submanifold of  $D$  with complex dimension  $(2\nu+1)N-1$ , holomorphic to the unit open ball of  $\mathbb{C}^{(2\nu+1)N-1}$ , and such that the polynomial  $\tilde{Q}(\vec{x})$  ( $\vec{x} \in D$ ) vanishes on  $\mathcal{N}$ . (Such a manifold  $\mathcal{N} \subset D$  exists by the induction hypothesis of Proposition 3.3.) We split the vectors  $\vec{x} \in D$  as  $\vec{x} = (\vec{y}, \vec{z})$ , so that the  $\vec{z}$ -part precisely contains the variables bearing the ball label  $N$ . We may assume that  $\mathcal{N} \subset B_1 \times B_2$ , where  $B_1$  and  $B_2$  are small, open balls in the spaces of the components  $\vec{y}$  and  $\vec{z}$ , respectively.

Assume, to the contrary of the statement of the sub-lemma, that the polynomial  $\tilde{Q}(\vec{x}) \equiv \tilde{Q}(\vec{y}, \vec{z})$  does depend on the component  $\vec{z}$ . Then, for typical but fixed values  $\vec{y}_0$  of  $\vec{y}$ , the “slice”  $\{\vec{y}_0\} \times B_2$  intersects the manifold  $\mathcal{N}$  in a set of complex codimension one. However, this fact clearly contradicts our earlier observation that the non-sufficiency of the orbit segments  $\omega \in D = D(\Sigma, \mathcal{A}, \vec{\tau} | m_N = 0)$  imposes a codimension-2 condition on the coordinates  $\vec{z}$  bearing the label  $N$ . This contradiction finishes the proof of the sub-lemma.  $\square$

**Finishing the proof of Proposition 3.3.** If  $Q(\vec{x})$  is the left-hand side of (1) or (2) in 3.3, then we arrive at the conclusion that the irreducible polynomial  $\tilde{Q}(\vec{x}) \equiv Q(\vec{x})$  divides  $\tilde{P}_1(\vec{x}), \dots, \tilde{P}_t(\vec{x})$ , and  $N \neq i_0$ ,  $N \neq j_0$  by Sub-lemma 3.8. This means, however, that the statement of the proposition is false for the  $(N-1)$ -ball system with the discrete algebraic scheme  $(\Sigma', \mathcal{A}', \vec{\tau}')$ , contradicting to our induction hypothesis.

If, however, the polynomial  $Q(\vec{x})$  is  $m_{i_0} + m_{j_0}$ , then in the case if  $N \notin \{i_0, j_0\}$  we arrive at a contradiction just the same way as above. If  $N \in \{i_0, j_0\}$ , say  $N = j_0$ , then  $\tilde{Q}(\vec{x}) \equiv m_{i_0}$ , and  $m_{i_0}$  is a common divisor of all polynomials  $\tilde{P}_1(\vec{x}), \dots, \tilde{P}_t(\vec{x})$  describing the non-sufficiency of the  $(\Sigma', \mathcal{A}', \vec{\tau}')$  subsystem with the  $N-1$  balls  $\{1, 2, \dots, N-1\}$ . This means that the above  $(\Sigma', \mathcal{A}', \vec{\tau}')$  subsystem is always non-sufficient, provided that  $m_{i_0} = 0$ . In the case  $N \geq 4$  it follows from Lemma 4.1 of [S-Sz(1999)] (applied to the  $(N-2)$ -ball system  $\{1, 2, \dots, N\} \setminus \{i_0, N\}$ ) and from the “non-equality” Lemma 3.6 that almost every  $(\Sigma', \mathcal{A}', \vec{\tau}')$ -orbit with  $m_{i_0} = 0$  is in fact sufficient. One easily checks by inspection that, in the case  $N = 3$ , actually every orbit of the 2-ball system  $\{1, 2\}$  with the side condition  $m_{i_0} = 0$  is hyperbolic (sufficient). The obtained contradiction finishes the inductive proof of Proposition 3.3.  $\square$

**Corollary 3.9.** Keep all the notations and assumptions of Proposition 3.3, except that we assume now that  $\Sigma = (\sigma_1, \dots, \sigma_n)$  is  $(2C(N) + 1)$ -rich and the singularity manifold  $C$  is defined by one of the following equations:

- (1)'  $\|v_{i_k}^k - v_{j_k}^k\|^2 = 0$ ,
- (2)'  $\langle v_{i_k}^k - v_{j_k}^k; \tilde{q}_{i_k}^k - \tilde{q}_{j_k}^k - L \cdot a_k \rangle = 0$ ,
- (3)'  $m_{i_k} + m_{j_k} = 0$

with some  $k$ ,  $1 \leq k \leq n$ . Let  $Q(\vec{x})$  be a common irreducible divisor of the polynomials  $P_1(\vec{x}), \dots, P_s(\vec{x})$  in (4.3) of [S-Sz(1999)] constructed for the entire discrete structure  $(\Sigma, \mathcal{A}, \vec{\tau})$  as above.

We again claim the same thing: The manifold  $C$  and the solution set of  $Q(\vec{x}) = 0$  cannot locally coincide. Consequently, an open, dense, and connected part of the irreducible variety  $V = \{Q(\vec{x}) = 0\}$  belongs to the domain  $D(\Sigma, \mathcal{A}, \vec{\tau})$  of the allowable initial data.

*Proof.* We write  $\Sigma = (\sigma_1, \dots, \sigma_n)$  in the form  $\Sigma = (\Sigma_1, \sigma_k, \Sigma_2)$ , where  $\Sigma_1 = (\sigma_1, \dots, \sigma_{k-1})$ ,  $\Sigma_2 = (\sigma_{k+1}, \dots, \sigma_n)$ . Then either  $\Sigma_1$  or  $\Sigma_2$  is  $C(N)$ -rich. If  $\Sigma_2$  turns out to be  $C(N)$ -rich, then we can directly apply the proposition after a simple time shift  $0 \mapsto k$ . In the other case, when we only know that  $\Sigma_1$  is  $C(N)$ -rich, beside the time shift  $0 \mapsto k$  an additional time-reversal is also necessary to facilitate the applicability of the proposition.  $\square$

Another consequence of Proposition 3.3 is

**Corollary 3.10.** Let  $\nu \geq 3$ ,  $(\Sigma, \mathcal{A}, \vec{\tau})$  be a discrete orbit structure with Property (A) and a  $C(N)$ -rich symbolic collision sequence  $\Sigma = (\sigma_1, \dots, \sigma_n)$ . Denote by  $P_1(\vec{x}), \dots, P_s(\vec{x})$  the polynomials of (4.3) of [S-Sz(1999)] just as before, and let  $Q(\vec{x})$  be a common irreducible divisor of  $P_1(\vec{x}), \dots, P_s(\vec{x})$ . (If such a common divisor exists.) Let, finally,  $(\bar{\Sigma}, \bar{\mathcal{A}}, \bar{\rho})$  be an extended discrete orbit structure with Property (A) and an extended collision sequence  $\bar{\Sigma} = (\sigma_0, \sigma_1, \dots, \sigma_n)$ . According to Lemma 3.1, we can find a particle, say the one with label  $N$ , and two indices  $1 \leq p < q \leq n$  such that

- (i)  $N \in \sigma_p \cap \sigma_q$ ,
- (ii)  $N \notin \bigcup_{j=p+1}^{q-1} \sigma_j$ ,
- (iii)  $\sigma_p = \sigma_q \implies (\exists j) (p < j < q \text{ \& } \sigma_p \cap \sigma_j \neq \emptyset)$ , and
- (iv)  $\Sigma'$  is  $(2C(N-1) + 1)$ -rich on the vertex set  $\{1, \dots, N-1\}$ .

(Here, just as in the case of derived schemes, we denote by  $\Sigma'$  the symbolic sequence that can be obtained from  $\Sigma$  by discarding all edges containing  $N$ .) Denote by  $\tilde{Q}(\vec{x})$  the polynomial that we obtain from  $Q(\vec{x})$  after the substitution  $m_N = 0$ . (Obviously,  $\tilde{Q}(\vec{x}) \not\equiv 0$ , otherwise there would not be any sufficient orbit segment  $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$  with  $m_N = 0$ .)

We claim that none of the irreducible polynomials on the left-hand side of

- (1)  $\|v_{i_0}^0 - v_{j_0}^0\|^2 = 0$ ,
- (2)  $\langle v_{i_0}^0 - v_{j_0}^0; \tilde{q}_{i_0}^0 - \tilde{q}_{j_0}^0 - L \cdot a_0 \rangle = 0$ ,
- (3)  $m_{i_0} + m_{j_0} = 0$

is a divisor of  $\tilde{Q}(\vec{x})$ .

*Proof.* Consider and fix an irreducible factor  $R(\vec{x})$  of  $\tilde{Q}(\vec{x})$ . According to Sublemma 3.8, the polynomial  $R(\vec{x}) \equiv R(\vec{y}, \vec{z})$  does not contain any variable bearing the label  $N$ , i.e.,  $R(\vec{x}) \equiv R(\vec{y})$ .

Assume, to the contrary of the statement that we want to prove, that the irreducible polynomial  $R(\vec{x}) \equiv R(\vec{y})$  is identical to one of the irreducible polynomials on the left-hand side of (1), (2), or (3). In particular, we have that  $N \neq i_0$ ,  $N \neq j_0$ . As we saw in the course of the proof of Sub-lemma 3.8, the algebraic variety  $V = \{\vec{y} \in \mathbb{C}^{(2\nu+1)(N-1)+1} \mid R(\vec{y}) = 0\}$ , defined by one of the equations (1), (2), or (3), describes the non-sufficiency of the derived system  $\Omega(\Sigma', \mathcal{A}', \vec{\tau}')$  that one obtains from the original  $\Omega(\Sigma, \mathcal{A}, \vec{\tau})$  by taking  $m_N = 0$ , i.e., for any point  $\vec{y} \in D(\Sigma', \mathcal{A}', \vec{\tau}')$  it is true that  $\vec{y} \in V$  if and only if there is some non-sufficient complex orbit segment  $\omega \in \Omega(\Sigma', \mathcal{A}', \vec{\tau}')$  with  $\vec{y}(\omega) = \vec{y}$ . However, this statement contradicts to the assertion of Proposition 3.3.  $\square$

**Remark.** Note that the polynomial  $\tilde{Q}(\vec{x})$  cannot be a constant  $c \neq 0$ , otherwise the original polynomial  $Q(\vec{x}) = c + m_N \cdot S(\vec{x})$  would not be homogeneous in the mass variables, see also the proof of Lemma 3.4.

The main result of this section is

**Key Lemma 3.11.** Keep all the notations and notions of this section. Assume that  $\nu \geq 3$  and the symbolic collision sequence  $\Sigma = (\sigma_1, \dots, \sigma_n)$  of the discrete algebraic frame  $(\Sigma, \mathcal{A}, \vec{\tau})$  (with Property (A)) is  $C(N)$ -rich.

We claim that all orbit segments  $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$  are sufficient apart from an algebraic variety of codimension-two (at least two, that is), i.e., the polynomials  $P_1(\vec{x}), \dots, P_s(\vec{x})$  of (4.3) of [S-Sz(1999)] do not have a non-constant common divisor.

*Proof.* The inductive proof employs many of the ideas of the proof of Proposition 3.3 and it will use the statement of the proposition itself. (More precisely, the statement of Corollary 3.9 is to be used.)

Indeed, the assertion of this key lemma is trivially true in the case  $N = 2$ , for in that case there are no non-sufficient, complex orbit segments  $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$  at all.

Assume that  $N \geq 3$ , and the statement of the key lemma has been successfully proven for all smaller values ( $2 \leq N' < N$ ). Our induction step is going to be a proof by contradiction. Suppose, therefore, that the polynomials  $P_1(\vec{x}), \dots, P_s(\vec{x})$  do have a common irreducible divisor  $Q(\vec{x})$ . Following the assertion of Lemma 3.1, select a suitable label  $k_0 \in \{1, \dots, N\}$  for the substitution  $m_{k_0} = 0$  so that a derived scheme  $(\Sigma', \mathcal{A}', \vec{\tau}')$  (with Property (A)) exists for the arising  $(N-1)$ -ball system  $\{1, \dots, N\} \setminus \{k_0\}$  with a symbolic sequence  $\Sigma'$ , possessing the properties (1)–(4) of Lemma 3.1, the same way as we did in the proof of Proposition 3.3. Without loss of generality, we may assume that  $k_0 = N$ .

Consider now the original system  $\Omega(\Sigma, \mathcal{A}, \vec{\tau})$  with the constraint  $m_N = 0$ . After the substitution  $m_N = 0$  the polynomial  $Q(\vec{x})$  becomes a new, non-constant polynomial  $\tilde{Q}(\vec{x})$ , see the proof of Sub-lemma 3.7 above. Let  $S(\vec{x})$  be an irreducible divisor of  $\tilde{Q}(\vec{x})$ . The (indecomposable) algebraic variety  $V = \{S(\vec{x}) = 0\}$  has one codimension, in the submanifold  $\tilde{\Omega} = \Omega_{m_N=0}$  of  $\Omega(\Sigma, \mathcal{A}, \vec{\tau})$ , and for every  $\vec{x} \in V$  there exists a non-hyperbolic complex orbit segment  $\omega \in \Omega(\Sigma, \mathcal{A}, \vec{\tau})$  with  $\vec{x}(\omega) = \vec{x}$

and  $m_N(\omega) = 0$ . As far as the non-sufficiency of the orbits  $\omega \in \Omega(\Sigma, \mathcal{A}, \bar{\tau})$  with  $m_N(\omega) = 0$  is concerned, we again take a close look at Lemma 4.9 of [S-Sz(1999)]. We saw earlier (see the proof of Lemma 3.6, which clearly carries over to the models subjected to the side condition  $m_N = 0$ ) that the parallelity of the relative velocities  $v_N^p(\omega) - v_{i_p}^p(\omega)$  and  $v_N^{q-1}(\omega) - v_{i_q}^{q-1}(\omega)$  takes place on a manifold with codimension at least two in our case of  $\nu \geq 3$ . Therefore, according to Lemma 4.9 of [S-Sz(1999)], the “codimension-one event”  $\bar{x}(\omega) \in V (\iff S(\bar{x}(\omega)) = 0)$  for orbits with  $m_N(\omega) = 0$  can only be equivalent to the non-sufficiency of the  $\{1, \dots, N-1\}$ -part  $\text{trunc}(\omega) \in \Omega(\Sigma', \mathcal{A}', \bar{\tau}')$  of the system. In this way it follows from Lemma 4.9 that the irreducible polynomial  $S(\bar{x})$  lacks all variables bearing the label  $N$ , see also the statement and the proof of Sub-lemma 3.8. We conclude that for every  $\bar{x} \in V$  (i.e., with  $S(\bar{x}) = 0$ ) there exists an orbit segment  $\omega \in \Omega(\Sigma, \mathcal{A}, \bar{\tau})$  with  $m_N(\omega) = 0$ ,  $\bar{x}(\omega) = \bar{x}$ , and a non-sufficient truncated segment  $\omega' = \text{trunc}(\omega) \in \Omega(\Sigma', \mathcal{A}', \bar{\tau}')$ . According to Proposition 3.3 above (applied to the  $(N-1)$ -ball system  $\{1, \dots, N-1\}$  with the algebraic scheme  $(\Sigma', \mathcal{A}', \bar{\tau}')$ ), the variety  $\{S(\bar{x}) = 0\}$  does not even locally coincide with the singularities of the complex dynamics  $\Omega(\Sigma', \mathcal{A}', \bar{\tau}')$ . This means that a codimension-one family of complex orbit segments  $\omega' = \text{trunc}(\omega) \in \Omega(\Sigma', \mathcal{A}', \bar{\tau}')$ ,  $\bar{x}(\omega) \in V$ , is not sufficient. This, in turn, contradicts the induction hypothesis of the proof of Key Lemma 3.11 by actually finishing it.  $\square$

## 4 Finishing the proof of ergodicity

### From $\mathbb{C}$ back to $\mathbb{R}$

First of all, we transfer the main result of the previous section (Key Lemma 3.11) from the complex set-up back to the real case. This result will be an almost immediate consequence of Key Lemma 3.11.

Fix a discrete algebraic scheme  $(\Sigma, \mathcal{A}, \bar{\tau})$  for  $N$  balls with Property (A) (see Definition 3.31 in [S-Sz(1999)]) and a  $C(N)$ -rich symbolic collision sequence  $\Sigma = (\sigma_1, \dots, \sigma_n)$ . (The definition of the threshold  $C(N)$  can be found in Lemma 3.1.) Denote by  $\Omega_{\mathbb{R}} = \Omega_{\mathbb{R}}(\Sigma, \mathcal{A}, \bar{\tau})$  the set of all elements  $\omega \in \Omega(\Sigma, \mathcal{A}, \bar{\tau})$  for which

- (1) all kinetic functions  $(\tilde{q}_i^k(\omega))_j$ ,  $(v_i^k(\omega))_j$ ,  $m_i(\omega)$ , and  $L(\omega)$  take real values,  $i = 1, \dots, N$ ;  $k = 0, 1, \dots, n$ ;  $j = 1, \dots, \nu$ ;
- (2)  $\tau_k(\omega) = t_k(\omega) - t_{k-1}(\omega) > 0$  for  $k = 1, \dots, n$ ;
- (3) out of the two real roots of (3.8) of [S-Sz(1999)] the root  $\tau_k$  is always selected as the smaller one,  $k = 1, \dots, n$ .

It is clear that either  $\Omega_{\mathbb{R}} = \Omega_{\mathbb{R}}(\Sigma, \mathcal{A}, \bar{\tau})$  is a  $((2\nu + 1)N + 1)$ -dimensional, real analytic submanifold of  $\Omega = \Omega(\Sigma, \mathcal{A}, \bar{\tau})$ , or  $\Omega_{\mathbb{R}} = \emptyset$ . Of course, we will never investigate the case  $\Omega_{\mathbb{R}} = \emptyset$ .

Consider the corresponding polynomials  $P_1(\bar{x}), \dots, P_s(\bar{x})$  of (4.3) of [S-Sz(1999)] describing the non-sufficiency of the complex orbit segments  $\omega \in \Omega(\Sigma, \mathcal{A}, \bar{\tau})$ , along the lines of Lemma 4.2 of [S-Sz(1999)], in terms of the kinetic

data  $\vec{x} = \vec{x}(\omega)$ . According to the statement in the third paragraph on p. 61 of [S-Sz(1999)], these polynomials  $P_i(\vec{x})$  admit real coefficients. By Key Lemma 3.11, the greatest common divisor of  $P_1(\vec{x}), \dots, P_s(\vec{x})$  is 1, hence the common zero set

$$\left\{ \vec{x} \in \mathbb{R}^{(2\nu+1)N+1} \mid P_1(\vec{x}) = P_2(\vec{x}) = \dots = P_s(\vec{x}) = 0 \right\}$$

of these polynomials does not contain any smooth real submanifold of (real) dimension  $(2\nu+1)N$ . In this way we obtained

**Proposition 4.1.** Use all the notions, notations and assumptions from above. There exists no smooth, real submanifold  $\mathcal{M}$  of  $\Omega_{\mathbb{R}}$  with  $\dim_{\mathbb{R}} \mathcal{M} = \dim_{\mathbb{R}} \Omega_{\mathbb{R}} - 1$  ( $= (2\nu+1)N$ ) and with the property that all orbit segments  $\omega \in \mathcal{M}$  are non-sufficient. (For the concept of non-sufficiency, please see §2.)  $\square$

### The “Fubini-type” argument

Our dynamics  $\Omega(\Sigma, \mathcal{A}, \vec{\tau})$  has the obvious feature that the variables  $m_i = m_i(\omega)$  ( $i = 1, \dots, N$ ) and  $L(\omega)$  (the so-called outer geometric parameters) remain unchanged during the time-evolution. Quite naturally, we do not need Proposition 4.1 directly but, rather, we need to use its analog for (almost) every fixed  $(N+1)$ -tuple  $(m_1, \dots, m_N; L) \in \mathbb{R}^{N+1}$ . This will be easily achieved by a classical “Fubini-type” product argument. The result is

**Proposition 4.2.** Use all the notions, notations and assumptions from above. Denote by

$$NS = NS(\Sigma, \mathcal{A}, \vec{\tau}) = \{ \omega \in \Omega_{\mathbb{R}}(\Sigma, \mathcal{A}, \vec{\tau}) \mid \dim_{\mathbb{C}} \mathcal{N}(\omega) > \nu + 1 \}$$

the set of all non-sufficient, real orbit segments  $\omega \in \Omega_{\mathbb{R}} = \Omega_{\mathbb{R}}(\Sigma, \mathcal{A}, \vec{\tau})$ . (For the definition of the complex neutral space  $\mathcal{N}(\omega)$ , please see (3.21) in [S-Sz(1999)].) Finally, we use the notation

$$\Omega_{\mathbb{R}}(\vec{m}, L) = \{ \omega \in \Omega_{\mathbb{R}} \mid \vec{m}(\omega) = \vec{m}, \text{ and } L(\omega) = L \}$$

for any given  $(N+1)$ -tuple  $(\vec{m}, L) = (m_1, \dots, m_N, L) \in \mathbb{R}^{N+1}$ . We claim that for almost every  $(\vec{m}, L) \in \mathbb{R}^{N+1}$  (for which  $\Omega_{\mathbb{R}}(\vec{m}, L) \neq \emptyset$ ) the intersection  $NS \cap \Omega_{\mathbb{R}}(\vec{m}, L)$  has at least 2 codimensions in  $\Omega_{\mathbb{R}}(\vec{m}, L)$ .

**Remark 4.3.** As it is always the case with such algebraic systems, the exceptional zero-measure set of the parameters  $(\vec{m}, L)$  turns out to be a countable union of smooth, proper submanifolds of  $\mathbb{R}^{N+1}$ .

**Proof of Proposition 4.2.** It is clear that the statement of the proposition is a local one, therefore it is enough to prove that for any small, open subset  $U_0 \subset \Omega_{\mathbb{R}}$  of  $\Omega_{\mathbb{R}} = \Omega_{\mathbb{R}}(\Sigma, \mathcal{A}, \vec{\tau})$  the set

$$\{ (\vec{m}, L) \in \mathbb{R}^{N+1} \mid \dim_{\mathbb{R}} (NS \cap \Omega_{\mathbb{R}}(\vec{m}, L) \cap U_0) \geq 2\nu N - 1 \}$$

of the “bad points”  $(\vec{m}, L)$  has zero Lebesgue measure. The points  $\omega \in U_0$  can be identified locally (in  $U_0$ ) with the vector  $\vec{x} = \vec{x}(\omega) \in D_{\mathbb{R}} = D(\Sigma, \mathcal{A}, \vec{\tau}) \cap \Omega_{\mathbb{R}}$  of their



initial coordinates. After this identification the small open set  $U_0 \subset \Omega_{\mathbb{R}}$  naturally becomes an open subset  $U_0 \subset D_{\mathbb{R}}$ . Furthermore, we split the points  $\vec{x} \in U_0$  as  $\vec{x} = ((\vec{m}, L), \vec{y})$ , where  $\vec{y}$  contains all variables other than  $m_1, \dots, m_N, L$ . In this way we may assume that  $U_0$  has a product structure  $U_0 = B_0 \times B_1$  of two small open balls, so that  $B_0 \subset \mathbb{R}^{N+1}$ , while the open ball  $B_1 \subset \mathbb{R}^{2\nu N}$  contains the  $\vec{y}$ -parts of the points  $\vec{x} = ((\vec{m}, L), \vec{y}) \in U_0$ .

Assume that the statement of the proposition is false. Then there exists a small open set  $U_0 = B_0 \times B_1 \subset \mathbb{R}^{N+1} \times \mathbb{R}^{2\nu N}$  (with the above splitting) and there is a positive number  $\epsilon_0$  such that the set

$$A_0 = \{(\vec{m}, L) \in B_0 \mid ((\vec{m}, L) \times B_1) \cap NS \text{ contains a } (2\nu N - 1)\text{-dimensional, smooth, real submanifold with inner radius } > \epsilon_0\}$$

has a positive Lebesgue measure in  $B_0$ . Then one can find an orthogonal projection  $P : \mathbb{R}^{2\nu N} \rightarrow H$  onto a hyperplane  $H$  of  $\mathbb{R}^{2\nu N}$  such that, by taking  $\Pi(\vec{x}) = \Pi((\vec{m}, L), \vec{y}) = P(\vec{y})$ ,  $(\Pi : \mathbb{R}^{(2\nu+1)N+1} \rightarrow H)$ , the set

$$A_1 = \{(\vec{m}, L) \in B_0 \mid \Pi[(\vec{m}, L) \times B_1] \cap NS \text{ contains an open ball of radius } > \epsilon_0/2 \text{ in } H\}$$

has positive Lebesgue measure in  $B_0$ . By the Fubini theorem the set

$$\tilde{\Pi}[NS \cap (B_0 \times B_1)]$$

has positive Lebesgue measure in  $B_0 \times H$ , where

$$\tilde{\Pi}(\vec{x}) = \tilde{\Pi}((\vec{m}, L), \vec{y}) = ((\vec{m}, L), P(\vec{y})) \in B_0 \times H$$

for  $\vec{x} \in B_0 \times B_1$ . However,  $\dim_{\mathbb{R}}(B_0 \times H) = (2\nu + 1)N$ , and  $\dim_{\mathbb{R}}(NS \cap \Omega_{\mathbb{R}}) \leq (2\nu + 1)N - 1$  (according to Proposition 4.1). Thus, we obtained that the algebraic set

$$\tilde{\Pi}(NS \cap (B_0 \times B_1)) \subset B_0 \times H$$

has dimension strictly less than  $\dim_{\mathbb{R}}(B_0 \times H) = (2\nu + 1)N$ , yet it has positive Lebesgue measure in the space  $B_0 \times H$ . The obtained contradiction finishes the proof of Proposition 4.2.  $\square$

### Finishing the proof of the theorem

We will carry out an induction with respect to the number of balls  $N (\geq 2)$ . For  $N = 2$  the system is well known to be a strictly dispersive billiard flow (after the obvious reductions  $m_1 v_1 + m_2 v_2 = 0$ ,  $m_1 \|v_1\|^2 + m_2 \|v_2\|^2 = 1$  ( $m_1, m_2 > 0$ ), and after the factorization with respect to the uniform spatial translations, as usual) and, as such, it is proved to be ergodic by Sinai in [Sin(1970)], see also the paper [S-W(1989)] about the case of different masses.

Assume now that  $N \geq 3$ ,  $\nu \geq 3$ , and the theorem has been successfully proven for all smaller numbers of balls  $N' < N$ . Suppose that a billiard flow

$$(\mathbf{M}, \{S^t\}_{t \in \mathbb{R}}, \mu) = (\mathbf{M}_{\vec{m}, L}, \{S_{\vec{m}, L}^t\}, \mu_{\vec{m}, L})$$

is given for  $N$  balls and outer geometric parameters  $(\vec{m}, L) = (m_1, \dots, m_N, L)$  ( $m_i > 0$ ,  $L > 0$ ) in such a way that, besides the always assumed properties (2.1.1)–(2.1.2),

- (\*) the vector  $(\vec{m}, L)$  of geometric parameters is such that for any subsystem  $1 \leq i_1 < i_2 < \dots < i_{N'} \leq N$  ( $2 \leq N' \leq N$ ) and for any  $C(N')$ -rich discrete algebraic scheme  $(\Sigma, \mathcal{A}, \vec{\tau})$  (with Property (A)), for this subsystem  $(m_{i_1}, \dots, m_{i_{N'}}, L)$  it is true that the parameter vector  $(m_{i_1}, \dots, m_{i_{N'}}, L)$  does not belong to the zero-measured exceptional set of parameters featuring Proposition 4.2.

According to Lemma 4.1 of [K-S-Sz(1990)], the set  $R_2 \subset \mathbf{M}$  of the phase points with at least two singularities on their trajectories is a countable union of smooth submanifolds of  $\mathbf{M}$  with codimension two, so this set  $R_2$  can be safely discarded in the proof, for it is slim, see also §2 about the slim sets. Secondly, by the induction hypothesis and by Theorem 5.1 of [Sim(1992-I)] (adapted to the case of different masses) there is a slim subset  $S_1 \subset \mathbf{M}$  such that for every phase point  $x \in \mathbf{M} \setminus S_1$

- (i)  $S^{(-\infty, \infty)}x$  contains at most one singularity, and
- (ii)  $S^{(-\infty, \infty)}x$  contains an arbitrarily large number of consecutive, connected collision graphs.

(In the case of a singular trajectory  $S^{(-\infty, \infty)}x$  we require that both branches contain an arbitrarily large number of consecutive, connected collision graphs.) Then, by Proposition 4.2 just proved, there is another slim subset  $S_2 \supset S_1$  of  $\mathbf{M}$  such that

- (H) for every  $x \in \mathbf{M} \setminus S_2$  the trajectory  $S^{(-\infty, \infty)}x$  contains at most one singularity and it is sufficient (or, geometrically hyperbolic).

According to Theorem 6.1 of [Sim(1992-I)] (easily adapted to the case of different masses) and Proposition 4.2, the so-called Chernov-Sinai Ansatz (see §2) holds true, i.e., for almost every singular phase point  $x \in \mathcal{SR}^+$  the positive semi-trajectory  $S^{(0, \infty)}x$  is non-singular and sufficient.

This is the point where the fundamental theorem for algebraic semi-dispersive billiards (Theorem 4.4 in [B-Ch-Sz-T(2002)]) comes to play! According to that theorem, by also using the crucial conditions (H) and the Ansatz above, it is true that for every phase point  $x \in (\text{int}\mathbf{M}) \setminus S_2$  some open neighborhood  $U_x$  of  $x$  in  $\mathbf{M}$  belongs to a single ergodic component of the considered billiard flow  $\{S_{\vec{m}, L}^t\}$ . Since the set  $(\text{int}\mathbf{M}) \setminus S_2$  contains an arc-wise connected set  $C$  with full  $\mu$ -measure (see Proposition 2.7.9 above), we get that the entire set  $C$  belongs to a single ergodic component of the flow  $\{S_{\vec{m}, L}^t\}$ . This finishes the proof of the ergodicity theorem.  $\square$

## 5 Concluding remark: The irrational mass ratio

Due to the natural reduction  $\sum_{i=1}^N m_i v_i = 0$  (which we always assume), in §§1–2 we had to factorize the configuration space with respect to spatial translations:  $(q_1, \dots, q_N) \sim (q_1 + a, \dots, q_N + a)$  for all  $a \in \mathbb{T}^\nu$ . It is a remarkable fact, however, that (despite the reduction  $\sum_{i=1}^N m_i v_i = 0$ ) even without this translation factorization the system still retains the Bernoulli mixing property, provided that the masses  $m_1, \dots, m_N$  are rationally independent. (We note that dropping the above-mentioned configuration factorization obviously introduces  $\nu$  zero Lyapunov exponents.) For the case  $N = 2$  (i.e., two disks) this was proven in [S-W(1989)] by successfully applying D. Rudolph's following theorem on the B-property of isometric group extensions of Bernoulli shifts [R(1978)]:

Suppose that we are given a dynamical system  $(M, T, \mu)$  with a probability measure  $\mu$  and an automorphism  $T$ . Assume that a compact metric group  $G$  is also given with the normalized Haar measure  $\lambda$  and left invariant metric  $\rho$ . Finally, let  $\varphi: M \rightarrow G$  be a measurable map. Consider the skew product dynamical system  $(M \times G, S, \mu \times \lambda)$  with  $S(x, g) = (Tx, \varphi(x) \cdot g)$ ,  $x \in M$ ,  $g \in G$ . We call the system  $(M \times G, S, \mu \times \lambda)$  an isometric group extension of the base (or factor)  $(M, T, \mu)$ . (The phrase “isometric” comes from the fact that the left translations  $\varphi(x) \cdot g$  are isometries of the group  $G$ .) Rudolph's mentioned theorem claims that the isometric group extension  $(M \times G, S, \mu \times \lambda)$  enjoys the B-mixing property as long as it is at least weakly mixing and the factor system  $(M, T, \mu)$  is a B-mixing system.

But how do we apply this theorem to show that the typical system of  $N$  hard balls in  $\mathbb{T}^\nu$  with  $\sum_{i=1}^N m_i v_i = 0$  is a Bernoulli flow, even if we do not make the factorization (of the configuration space) with respect to uniform spatial translations? It is simple. The base system  $(M, T, \mu)$  of the isometric group extension  $(M \times G, S, \mu \times \lambda)$  will be the time-one map of the factorized (with respect to spatial translations) hard ball system. The group  $G$  will be just the container torus  $\mathbb{T}^\nu$  with its standard Euklidean metric  $\rho$  and normalized Haar measure  $\lambda$ . The second component  $g$  of a phase point  $y = (x, g) \in M \times G$  will be just the position of the center of the (say) first ball in  $\mathbb{T}^\nu$ . Finally, the governing translation  $\varphi(x) \in \mathbb{T}^\nu$  is quite naturally the total displacement

$$\int_0^1 v_1(x_t) dt \quad (\text{mod } \mathbb{Z}^\nu)$$

of the first particle while unity of time elapses. In the previous sections the B-mixing property of the factor map  $(M, T, \mu)$  has been proven successfully for typical geometric parameters  $(m_1, \dots, m_N; L)$ . Then the key step in proving the B-property of the isometric group extension  $(M \times G, S, \mu \times \lambda)$  is to show that the latter system is weakly mixing. This is just the essential contents of the article [S-W(1989)], and it takes advantage of the assumption of rational independence of the masses. Here we are only presenting to the reader the outline of that proof. As a matter of fact, we not only proved the weak mixing property of the extension

$(M \times G, S, \mu \times \lambda)$ , but we showed that this system has in fact the K-mixing property by proving that the Pinsker partition  $\pi$  of  $(M \times G, S, \mu \times \lambda)$  is trivial. (The Pinsker partition is, by definition, the finest invariant, measurable partition of the dynamical system with respect to which the factor system has zero metric entropy. A dynamical system is K-mixing if and only if its Pinsker partition is trivial, i.e., it consists of only the sets with measure zero and one, see [K-S-F(1980)].) In order to show that the Pinsker partition is trivial, in [S-W(1989)] we constructed a pair of measurable partitions  $(\xi^s, \xi^u)$  for  $(M \times G, S, \mu \times \lambda)$  made up by open and connected sub-manifolds of the local stable and unstable manifolds, respectively. It followed by standard methods (see [Sin(1968)]) that the partition  $\pi$  is coarser than each of  $\xi^s$  and  $\xi^u$ . Due to the  $S$ -invariance of  $\pi$ , we have that  $\pi$  is coarser than

$$\bigwedge_{n \in \mathbb{Z}} S^n \xi^s \wedge \bigwedge_{n \in \mathbb{Z}} S^n \xi^u. \quad (*)$$

In the final step, by using now the rational independence of the masses, we showed that the partition in  $(*)$  is, indeed, trivial.

## References

- [B-Ch-Sz-T(2002)] P. Bálint, N. Chernov, D. Szász, I.P. Tóth, Multidimensional semidispersing billiards: singularities and the fundamental theorem, *Ann. Henri Poincaré* **3**, No. 3, 451–482 (2002).
- [B-F-K(1998)] D. Burago, S. Ferleger, A. Kononenko, Uniform estimates on the number of collisions in semi-dispersing billiards, *Annals of Mathematics* **147**, 695–708 (1998).
- [B-L-P-S(1992)] L. Bunimovich, C. Liverani, A. Pellegrinotti, Yu. Sukhov, Special Systems of Hard Balls that Are Ergodic, *Commun. Math. Phys.* **146**, 357–396 (1992).
- [B-S(1973)] L.A. Bunimovich, Ya.G. Sinai, The fundamental theorem of the theory of scattering billiards, *Math. USSR-Sb.* **19**, 407–423 (1973).
- [C-H(1996)] N.I. Chernov, C. Haskell, Nonuniformly hyperbolic K-systems are Bernoulli, *Ergod. Th. & Dynam. Sys.* **16**, 19–44 (1996).
- [E(1978)] R. Engelking, Dimension Theory, *North Holland* (1978).
- [G(1981)] G. Galperin, On systems of locally interacting and repelling particles moving in space, *Trudy MMO* **43**, 142–196 (1981).
- [K(1942)] N.S. Krylov, The Processes of Relaxation of Statistical Systems and the Criterion of Mechanical Instability, Thesis, Moscow, (1942);

Republished in English by Princeton University Press, Princeton N.J. (1979).

- [K-S(1986)]      A. Katok, J.-M. Strelcyn, Invariant Manifolds, Entropy and Billiards; Smooth Maps with Singularities, *Lecture Notes in Mathematics* **1222**, Springer (1986).
- [K-S-F(1980)]    I.P. Kornfeld, Ya.G. Sinai, S.V. Fomin, Ergodic Theory, Nauka, Moscow (1980).
- [K-S-Sz(1989)]   A. Krámli, N. Simányi, D. Szász, Ergodic Properties of Semi-Dispersing Billiards I. Two Cylindric Scatterers in the 3-D Torus, *Nonlinearity* **2**, 311–326 (1989).
- [K-S-Sz(1990)]   A. Krámli, N. Simányi, D. Szász, A “Transversal” Fundamental Theorem for Semi-Dispersing Billiards, *Commun. Math. Phys.* **129**, 535–560 (1990).
- [K-S-Sz(1991)]   A. Krámli, N. Simányi, D. Szász, The K-Property of Three Billiard Balls, *Annals of Mathematics* **133**, 37–72 (1991).
- [K-S-Sz(1992)]   A. Krámli, N. Simányi, D. Szász, The K-Property of Four Billiard Balls, *Commun. Math. Phys.* **144**, 107–148 (1992).
- [L-W(1995)]      C. Liverani, M. Wojtkowski, Ergodicity in Hamiltonian systems, *Dynamics Reported* **4**, 130–202 (1995), arXiv:math.DS/9210229.
- [M(1976)]        D. Mumford, Algebraic Geometry I. Complex Projective Varieties, Springer Verlag, Berlin Heidelberg (1976).
- [O-W(1998)]      D. Ornstein, B. Weiss, On the Bernoulli Nature of Systems with Some Hyperbolic Structure, *Ergod. Th. & Dynam. Sys.* **18**, 441–456 (1998).
- [P(1977)]        Ya. Pesin, Characteristic Exponents and Smooth Ergodic Theory, *Russian Math. surveys* **32**, 55–114 (1977).
- [R(1978)]        D.J. Rudolph, Classifying the isometric extensions of a Bernoulli shift, *J. d’Anal. Math.* **34**, 36–50 (1978).
- [Sim(1992)-I]    N. Simányi, The K-property of  $N$  billiard balls I, *Invent. Math.* **108**, 521–548 (1992).
- [Sim(1992)-II]   N. Simányi, The K-property of  $N$  billiard balls II, *Invent. Math.* **110**, 151–172 (1992).

- [Sim(2003)]     N. Simányi, Proof of the Boltzmann-Sinai Ergodic Hypothesis for Typical Hard Disk Systems, *Inventiones Mathematicae* **154** No. 1, 123–178 (2003).
- [Sin(1963)]     Ya.G. Sinai, On the Foundation of the Ergodic Hypothesis for a Dynamical System of Statistical Mechanics, *Soviet Math. Dokl.* **4**, 1818–1822 (1963).
- [Sin(1968)]     Ya.G. Sinai, Dynamical systems with countably multiple Lebesgue spectrum II, *Amer. Math. Soc. Transl.* **68** No. 2, 34–38 (1968).
- [Sin(1970)]     Ya.G. Sinai, Dynamical Systems with Elastic Reflections, *Russian Math. Surveys* **25**, 2, 137–189 (1970).
- [St(1973)]     I. Stewart, Galois Theory, Chapman and Hall, London (1973).
- [S-Ch(1987)]     Ya.G. Sinai, N.I. Chernov, Ergodic properties of certain systems of 2-D discs and 3-D balls, *Russian Math. Surveys* **42** No. 3, 181–207 (1987).
- [S-Sz(1995)]     N. Simányi, D. Szász, The K-property of Hamiltonian systems with restricted hard ball interactions, *Mathematical Research Letters* **2** No. 6, 751–770 (1995).
- [S-Sz(1999)]     N. Simányi, D. Szász, Hard ball systems are completely hyperbolic, *Annals of Mathematics* **149**, 35–96 (1999).
- [S-Sz(2000)]     N. Simányi, D. Szász, Non-integrability of Cylindric Billiards and Transitive Lie Group Actions, *Ergod. Th. & Dynam. Sys.* **20**, 593–610 (2000).
- [S-W(1989)]     N. Simányi, M. Wojtkowski, Two-particle billiard system with arbitrary mass ratio, *Ergod. Th. & Dynam. Sys.* **9**, 165–171 (1989).
- [Sz(1994)]     D. Szász, The K-property of ‘Orthogonal’ Cylindric Billiards, *Commun. Math. Phys.* **160**, 581–597 (1994).
- [Sz(1996)]     D. Szász, Boltzmann’s Ergodic Hypothesis, a Conjecture for Centuries, *Studia Sci. Math. Hung* **31**, 299–322 (1996).
- [V(1979)]     L.N. Vaserstein, On Systems of Particles with Finite Range and/or Repulsive Interactions, *Commun. Math. Phys.* **69**, 31–56 (1979).
- [VDW(1970)]     B.L. van der Waerden, Algebra I, Frederick Ungar Publ. Co. (1970).

- [W(1988)]            M. Wojtkowski, Measure theoretic entropy of the system of hard spheres, *Ergod. Th. & Dynam. Sys.* **8**, 133–153 (1988).
- [W(1990)]            M. Wojtkowski, Linearly stable orbits in 3-dimensional billiards, *Commun. Math. Phys.* **129** No. 2, 319–327 (1990).

Nándor Simányi  
University of Alabama at Birmingham  
Department of Mathematics  
Campbell Hall  
Birmingham, AL 35294  
USA  
email: simanyimath.uab.edu

Communicated by Eduard Zehnder  
submitted 17/10/02, accepted 01/12/03



To access this journal online:  
<http://www.birkhauser.ch>

---